

# **Liquidity and Prices on the Real Estate Market: Determinants, Dynamics and Market Segmentation**



Dissertation zur Erlangung des Grades eines  
Doktors der Wirtschaftswissenschaft

eingereicht an der Fakultät für Wirtschaftswissenschaften  
der Universität Regensburg

Vorgelegt von:

Anna Freudenreich

Tag der Disputation:

27. November 2024

Berichterstatter:

Prof. Dr. Wolfgang Schäfers

Prof. Dr. Tobias Just



## Table of Contents

<b>List of Figures</b> .....	<b>III</b>
<b>List of Tables</b> .....	<b>IV</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Motivation and background.....	1
1.2 Research questions.....	4
1.3 Co-authors, submissions and conference presentations.....	5
1.4 References .....	7
<b>2 Liquidity and prices: A cluster analysis of the German residential real estate market</b> <b>9</b>	
2.1 Introduction .....	10
2.2 Literature .....	13
2.3 Data and descriptive statistics.....	16
2.4 Econometric approach .....	23
2.4.1 The residential price index .....	24
2.4.2 The residential liquidity index.....	25
2.4.3 Cluster analysis .....	26
2.5 Regional cluster results.....	28
2.5.1 Cluster results on the investment market.....	28
2.5.2 Cluster results on the rental market .....	35
2.5.3 Summary and further implications .....	41
2.6 Conclusion .....	43
2.7 References .....	46
<b>3 Exploring the determinants of real estate liquidity from an alternative perspective: censored quantile regression in real estate research</b> .....	<b>50</b>
3.1 Introduction .....	51
3.2 Literature review.....	54
3.3 Econometric approach .....	56
3.3.1 Cox proportional hazards model .....	56

3.3.2	Cox proportional hazards model versus censored quantile regression .....	57
3.3.3	Quantile regression model .....	57
3.3.4	Censored quantile regression model .....	58
3.4	Data and descriptive statistics .....	59
3.5	Results .....	65
3.5.1	Results of the cox survival regression .....	65
3.5.2	Results of the censored quantile regression .....	68
3.6	Conclusion .....	78
3.7	References .....	81
<b>4</b>	<b>What are tenants demanding the most? A machine learning approach for the prediction of time on market .....</b>	<b>84</b>
4.1	Introduction .....	85
4.2	Literature review .....	86
4.3	Econometric approach .....	90
4.4	Data and descriptive statistics .....	92
4.5	Results .....	94
4.5.1	Variable importance .....	94
4.5.2	Results of the random survival forest model .....	95
4.6	Conclusion .....	99
4.7	References .....	101
4.8	Appendix .....	105
<b>5</b>	<b>Conclusion .....</b>	<b>106</b>
5.1	Executive summaries .....	106
5.2	Final remarks .....	111
5.3	References .....	113

## List of Figures

<b>Figure 2.1</b> Generation process of residential real estate data .....	17
<b>Figure 2.2</b> Cumulative percentage change in mean price and mean time on market .....	20
<b>Figure 2.3</b> Distribution of selected variables across purchasing power percentiles .....	22
<b>Figure 2.4</b> Price index, investment market.....	28
<b>Figure 2.5</b> Liquidity index, investment market .....	30
<b>Figure 2.6</b> Socioeconomic data by cluster affiliation, investment market .....	32
<b>Figure 2.7</b> Allocation of NUTS 3 regions to price and liquidity clusters, investment market ....	34
<b>Figure 2.8</b> Price index, rental market .....	36
<b>Figure 2.9</b> Liquidity index, rental market .....	37
<b>Figure 2.10</b> Socioeconomic data by cluster affiliation, rental market.....	39
<b>Figure 2.11</b> Allocation of NUTS3 regions to price and liquidity clusters, rental market .....	40
<b>Figure 3.1</b> Average time on market in weeks in the seven largest German cities.....	52
<b>Figure 3.2</b> Descriptive development of selected covariates for time on market quantiles .....	64
<b>Figure 3.3</b> Estimated quantile regression coefficients of the quarterly dummy variables .....	70
<b>Figure 3.4</b> Estimated quantile regression coefficients of selected housing variables .....	73
<b>Figure 3.5</b> Estimated quantile regression coefficients of the spatial gravity variable .....	75
<b>Figure 3.6</b> Estimated quantile regression coefficients of the socioeconomic variables .....	77
<b>Figure 4.1</b> Partial dependency plots for the ten most influential variables on predicted survival time .....	98

## List of Tables

<b>Table 2.1</b> Variables and sources .....	19
<b>Table 3.1</b> Descriptive statistics of selected variables .....	61
<b>Table 3.2</b> Degree of censoring and average time on market in weeks per quantile.....	63
<b>Table 3.3</b> Results of the Cox survival regression .....	67
<b>Table 4.1</b> Descriptive statistics .....	94
<b>Table 4.2</b> Variable importance of 10 most influential variables on predicted survival time .....	95
<b>Table 4.3</b> Non-geographic features - dwelling characteristics.....	105

## 1 Introduction

### Motivation and background

Since the recent financial crisis, the German economy has experienced a strong and enduring expansion. The fundamental economic data exhibits a growing GDP, an increase in household incomes, high levels of labor demand and strong purchasing power. The consistently favorable macroeconomic situation and geographical events triggered high migration to Germany. Additionally, the number of households has been increasing, due to the social trend towards a smaller household size. Furthermore, interest rates for mortgages have been extremely low. Unsurprisingly, this economic development led to a booming demand for residential real estate, facing way too little housing supply especially in the metropolises. Consequently, vacancy has mostly been diminishing and housing prices as well as rents have been increasing massively. Despite rising construction activity, building completions have been insufficient to meet demand in several regions. Thus, especially in the urban areas, prices and rents have experienced enormous upside pressure. With a homeownership rate of 43% as of 2013 (GdW 2016), the first year analyzed in this dissertation, Germany shows to be a nation of tenants. Hence, the residential rental real estate market deserves closer attention. Voigtlaender (2009), Bentzien et al. (2012), Lerbs and Oberst (2014), Kohl (2016) and Reisenbichler (2016), among others, discuss in detail the reasons for this distinctive market feature. However, research on the nationwide rental market is rather scarce. Another peculiarity of the German market is the high percentage of private landlords. According to the GdW about 3.9 million non-professional private landlords offer about 37% of all apartments, while only 20% are offered by professional landlords. Currently, the real estate market assessment is mainly based on the evaluation of prices. However, on the residential real estate market the process of marketing a dwelling comprises of two essential components. The marketing process starts with the introduction of the dwelling onto the market at a specific asking price, which is the first component, determined by the seller or landlord. The second integral component is the time a dwelling is listed on the market until a buyer or tenant commits to pay the price and take the dwelling off the market. The easier, thus faster, the matching occurs, the shorter the time a dwelling is on the market and thus the higher its liquidity. In this dissertation the liquidity of a dwelling is defined as the inverse of its time on market. During times of increasing prices and thus decreasing returns of residential apartments it is crucial for sellers and landlords to sell or rent out their apartments as quickly as possible, i.e. to keep the time on market short. Consequently, the target variable is the time an apartment is up for sale or rent on the market. Hence, a duration is estimated. Therefore, another modelling class named survival models is needed. The most popular approach to analyze the time on market is the Cox (1972) proportional hazards model (PHM). In the field of real estate Kluger and Miller (1990) were the first to apply the semi-parametric Cox PHM. Several articles like Krainer (1990), Smith (2010), Hoeberichts et al. (2013) and Cirman et al. (2015) have applied this approach to estimate the

marketing time during the last years. Cajias and Freudenreich (2018) also apply the Cox PHM for the analysis of time on market of residential rental apartments in the seven largest German cities.

The aim of this dissertation is to analyze the second insightful and under-researched component in the process of marketing a residential dwelling, namely the time on market. The regional focus is on the German residential real estate market. The process starts with examining the whole German residential investment and rental market in Paper 1, continues with regarding the rental market of the seven largest German cities in Paper 2 and terminates with the in-depth analysis of the rental market of a single German city in Paper 3. Each of the three papers analyses the time on market from another methodological perspective and hence the course of this dissertation displays a methodological advancement.

The purpose of the first paper is to provide deeper insights to the highly under-researched German residential real estate market by a combined analysis of price and liquidity indices. Due to the strong heterogeneity within Germany, it is necessary to investigate the German residential market on a regional basis. As the rental market is the predominant residential market in Germany it is crucial to analyse not only the investment but also the rental market. Therefore, quality- and spatial-adjusted price and liquidity indices are calculated separately for the investment and rental market on a regional basis. The dataset underlying this analysis comprises more than 4.5 million observations in 380 German regions from 2013 Q1 to 2018 Q4. The regions are clustered with respect to their price and liquidity development. The clusters are then analyzed by means of further economic and socioeconomic data to identify similarities. Furthermore, the clusters are interpreted from a geographic perspective. The results show that the allocation to cluster 1 is always supported by higher growth rates in the variables, population, working population and real GDP, implying higher demand for space. Moreover, in each of the analyzed categories, cluster 1 reveals a lower unemployment rate as well as a higher disposable income. One of the most interesting implications is, that apparently a large part of the German population has developed into professional real estate investors. In Germany the largest share of landlords is the one of the so-called non-professional landlords. As the regions assigned to cluster 1, displaying the most significant price increase, seem to be chosen based on a very sophisticated market analysis by identifying the regions with the strongest fundamental data, it seems like the dominating market players have significantly increased their knowledge and approach for investing in residential real estate.

The aim of the second paper is to derive a detailed understanding of the major drivers of marketing time of rental dwellings. Therefore, the liquidity (inverse of time on market) of rental dwellings and its determinants for different liquidity quantiles for the seven largest German cities are examined. The article introduces a modelling technique which is new to the field and uses its results to validate the outcome of the Cox (1972) PHM which is the most popular approach measuring the time on market. The determinants are estimated using censored quantile

regressions (CQR) to investigate the impact on very liquid to very illiquid dwellings. Hence, applying a CQR, the impact of a covariate on the dependent variable time on market can be quantified for any quantile of the distribution. According to Chaudhuri et al. (1997) this is a great advantage compared to mean regressions as distributions might not only differ in terms of their means but especially at their upper and lower tails. As market heterogeneity is not only observed between cities but also within a city, each of the seven cities is considered individually. Micro data for almost 500,000 observations from 2013 to 2017 is used. Substantial differences in the magnitude and direction of the regression coefficients for the different liquidity quantiles are found. Furthermore, both the magnitude and direction of the impact of an explanatory variable on the liquidity, differ between the cities. To the best of the authors' knowledge this is the first paper, to apply censored quantile regressions to liquidity analysis of the real estate rental market.

The aim of the third paper is to detect the most important determinants as well as their influence on the time on market of residential rental apartments in the city of Munich. Therefore, a machine learning approach is introduced to the analysis of time on market on the real estate market. Although machine learning techniques have been applied frequently on the real estate market for the analysis of prices, its use for examining time on market is completely novel. The article empirically analyzes the most influential variables that affect the time on market of rental apartments for the city of Munich. Therefore, the random forest machine learning technique based on decision trees is applied. To the best of the authors' knowledge this is the first paper, to apply the random survival forest approach to the liquidity analysis on the real estate market. Micro data for 115,264 observations on the residential rental market of Munich from 2013 to 2021 is used. As a first step, the most important housing, social and spatial predictors of time on market on the residential rental market are revealed. Second, this paper analyses how the survival probability of residential rental apartments responds to these major characteristics based on a random survival forest approach. Hence, the segmentation of dwellings generated by the decision tree methodology results in a deep and profound understanding of the driving factors of liquidity.

In summary, the insights of this dissertation are interesting for further research as well as for practical implications. A profound understanding of the time on market on the residential real estate market in Germany is of great importance for private, institutional, and governmental market players. It is essential in order to evaluate market fundamentals, to assess and forecast cyclical movements as well as to derive policy implications. With a deep understanding of the time on market, its determinants and its behavior to changes in its determinants, real estate professionals as well as private investors are able to derive investment decisions and implications for portfolio management strategies and ultimately to minimize cash-flow failure.

## **Research questions**

In this section the essential research questions are presented individually for each paper. The core of this dissertation is the analysis of the time on market of residential dwellings on the German real estate market. Each of the three papers analyses the time on market from another methodological perspective.

### **Paper 1: Liquidity and prices: A cluster analysis of the German residential real estate market**

---

- 1) How did price and liquidity on the German residential investment and rental market measured by quality- and spatial-adjusted hedonic indices evolve over the last 6 years?
- 2) To what extent can the markets be clustered? What similarities and differences in view of economic and socioeconomic indicators do these regions share?
- 3) Which overall conclusions and implications can be drawn for the German residential real estate market?

### **Paper 2: Exploring the determinants of real estate liquidity from an alternative perspective: censored quantile regression in real estate research**

---

- 1) Do the direction and magnitude of the effect individual covariates have on residential real estate liquidity vary along the liquidity distribution, i.e. across low, medium, and high-liquidity segments?
- 2) Can the assumption of proportional hazards underlying the Cox PHM be confirmed or is there a need for an approach able to deal with heterogeneous effects?
- 3) Do the direction and magnitude of the effect individual covariates have on residential real estate liquidity vary across the seven cities?
- 4) What information about the marketability of their current and future apartments can landlords draw from the CQR results and how can they improve the marketability?

**Paper 3: What are tenants demanding the most? A machine learning approach for the prediction of time on market**

---

- 1) What are the most important housing, social and spatial characteristics affecting the survival probability of residential rental dwellings in Munich?
- 2) How does the survival probability of residential rental dwellings respond to these major characteristics based on a random survival forest approach?

**Co-authors, submissions and conference presentations**

The following overview provides information about co-authors, journal submissions, publication status and conference presentations.

**Paper 1: Liquidity and prices: A cluster analysis of the German residential real estate market**

---

**Authors:**

- Dr. Marcelo Cajias, Anna Freudenreich, Dr. Philipp Freudenreich, Prof. Dr. Wolfgang Schäfers

**Submission Details:**

- Journal: Journal of Business Economics
- Current Status: accepted and published online (16/08/2020)

**Conference Presentations:**

- This paper was presented at the 2018 Annual Conference of the European Real Estate Society (ERES) in Reading, United Kingdom.

**Paper 2: Exploring the determinants of real estate liquidity from an alternative perspective: censored quantile regression in real estate research**

---

**Authors:**

- Anna Freudenreich, Dr. Philipp Freudenreich, Dr. Marcelo Cajias

**Submission Details:**

- Journal: Journal of Business Economics
- Current Status: accepted and published online (03/06/2020)

**Conference Presentations:**

- This paper was presented at the 2018 Annual Conference of the ERES in Reading, United Kingdom.

**Paper 3: What are tenants demanding the most? A machine learning approach for the prediction of time on market**

---

**Authors:**

- Dr. Marcelo Cajias, Anna Freudenreich

**Submission Details:**

- Journal: Journal of Property Investment & Finance
- Current Status: accepted and published (13/02/2024)

**Conference Presentations:**

- This paper was presented at the 2023 Annual Conference of the European Real Estate Society (ERES) in London, United Kingdom.

---

**References**

- Bentzien, V., Rottke, N. and Zietz, J. (2012). Affordability and Germany's low homeownership rate. *International Journal of Housing Markets and Analysis*, 5(3), 289-312.
- Cajias, M. and Freudenreich, P. (2018). Exploring the determinants of liquidity with big data - market heterogeneity in German markets. *Journal of Property Investment and Finance*, 36(1), 3-18.
- Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics*, 25(2), 715-744.
- Cirman, A., Pahor, M. and Verbic, M. (2015). Determinants of time on the market in a thin real estate market. *Engineering Economics*, 26(1), 4-11.
- Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.
- GdW Bundesverband deutscher Wohnungs- und Immobilienunternehmen e.V. (2016). Anbieterstruktur auf dem deutschen Wohnungsmarkt. <https://www.gdw.de/uploads/pdf/infografiken/2016/Anbieterstruktur.pdf>. Accessed 01 February 2024
- Hoerberichts, M., van Rooij, M. and Siegmann, A. (2013). House list prices and durations in boom and bust. <https://deliverypdf.ssrn.com/delivery.php?ID=843106086008086070090023068097075010017051024001008020124117114123013031071031114091049101001060018038058112064065030069080124109034011046042124019015110117120004041055094020004081108098014124083003088083001095105009068120096120006097119117122098099&EXT=pdf&INDEX=TRUE>. Accessed 01 February 2024.
- Kluger, B.D. and Miller, N.G. (1990). Measuring residential real estate liquidity. *Real Estate Economics*, 18(2), 145-159.
- Kohl, S. (2016). Urban history matters: Explaining the German-American homeownership gap. *Housing Studies*, 31(6), 694-713.
- Krainer, J. (1999). Real estate liquidity. *Economic Review - Federal Reserve Bank of San Francisco*, 3, 14-26.
- Lerbs, O. and Oberst, A. (2014). Explaining the spatial variation in homeownership rates: Results for German regions. *Regional Studies*, 48(5), 844-865.
- Reisenbichler, A. (2016). A rocky path to homeownership: Why Germany eliminated large-scale subsidies for homeowners. *Cityscape*, 18(3), 283-290.
- Smith, B.C. (2010). Spatial Heterogeneity in Listing Duration: The Influence of Relative Location to Marketability. *Journal of Housing Research*, 18(2), 151-171.

Voigtlaender, M. (2009). Why is the German homeownership rate so low? *Housing Studies*, 24(3), 335-372.

## 2 Liquidity and prices: A cluster analysis of the German residential real estate market

### Abstract

This paper analyses the highly under-researched German residential real estate market. Quality- and spatial-adjusted price and liquidity indices are calculated separately for the investment and rental market on a regional basis. Applying the “Partitioning Around Medoids (PAM)” clustering algorithm, the regions are clustered with respect to their price and liquidity development after the average silhouette method is applied to find the optimal number of clusters. The dataset underlying this analysis comprises more than 4.5 million observations in 380 German regions from 2013 Q1 to 2018 Q4. The clusters are then analysed by means of further economic and socioeconomic data in order to identify similarities. Furthermore, the clusters are interpreted from a geographic perspective. We find that the allocation to cluster 1 is always supported by higher growth rates in the variables, population, working population and real GDP, implying higher demand for space. Moreover, in each of the analysed categories cluster 1 reveals a lower unemployment rate as well as a higher disposable income. One of the most interesting implications is, that apparently a large part of the German population has developed into professional real estate investors. In Germany the largest share of landlords is the one of the so-called non-professional landlords. As the regions assigned to cluster 1, displaying the most significant price increase, seem to be chosen based on a very sophisticated market analysis by identifying the regions with the strongest fundamental data, it seems like the dominating market players have significantly increased their knowledge and approach for investing in residential real estate.

**Acknowledgement:** The authors especially thank PATRIZIA Immobilien AG for contributing the dataset for this study.

All statements of opinion are those of the authors and do not necessarily reflect the opinions of PATRIZIA Immobilien AG or its associated companies.

## 2.1 Introduction

Financial assets such as stocks and bonds are traded in tremendous volumes, turning over billions of dollars within seconds and with almost no spatial constraints. By contrast, the transaction process of direct real estate is more complex, often consuming several months due to the heterogeneity of individual properties and market specific frictions. For example, larger participation-, search- and transaction-costs, as well as considerable asymmetric information impede a smooth match between buyer's or tenant's and seller's or landlord's price expectation within "short" time intervals. When it comes to residential real estate – an asset class which is strongly linked to individual preferences of buyers and tenants as well as expectations of sellers and landlords – a matching of both sides may be even more difficult. Thereby, the location of the respective dwelling plays a major role for the development of the key features determining the matching process.

On the residential real estate market, the process of selling or renting out a dwelling comprises of two essential components. The first component is the introduction of the dwelling onto the market at a price<sup>1</sup> determined by the seller or landlord. The second component is the time it takes until a prospective buyer or tenant is willing to take the dwelling off the market and to pay the price.<sup>2</sup> Contingent upon a matching of the expectations of supply and demand, a market is able to operate. The easier, thus faster this matching occurs, the higher the liquidity on the market. Liquidity is related to price in both cause and effect. According to Geltner et al. (2014), increased market activity, displayed by a higher number of transactions as well as a rising transaction frequency on the residential market, leads to more similar properties in the respective market. Hence, the observation of relevant transaction prices becomes easier. As the information of properties being transacted in the relevant market directly influences the individuals' price assessment of other properties, potential market participants become better informed, leading to less market uncertainty. Thus, potential buyers or tenants and sellers or landlords can form improved reservation prices, so that the respective distributions become narrower and converge to the "equilibrium price"<sup>3</sup>. Consequently, transaction prices improve and so does liquidity. In the following, liquidity is defined as the inverse of the time on market (TOM). Typically, the matching will occur faster if the price of the dwelling is supported by its particular location and building characteristics. Depending on the level of demand, buyers and tenants might start to accept higher prices. But as long as there is sufficient supply, the prospective buyer or tenant will continue to search the market and not rush into an undesired contractual agreement. Therefore,

---

<sup>1</sup> In the following, the term "price" refers to the price on the investment as well as the rental market and is distinguished as "price on the investment market" and "price on the rental market".

<sup>2</sup> Due to a lack of transaction prices and contract rents on the German residential market the asking price and asking rent, respectively, operate as a "take it or leave it option" to the buyer or tenant and thus price negotiations are not considered as well. (see section "2.3. Data and Descriptive Statistics")

<sup>3</sup> The "equilibrium price" is defined as the price, where the market clears, leading to maximum transaction volume and maximum liquidity. This is the case, when there is no asset price uncertainty and all potential market participants are perfectly informed.

the buyer or tenant is said to be the provider of liquidity, as he has the financial resources to afford the dwelling and to convert it into cash or a dividend yielding asset for the owner or landlord, see Fisher et al. (2003). Only if it is up to “take what you can get”, buyers and tenants will be accepting a price which is exceeding their initial reservation price in no time.

Currently, the assessment of the real estate market is mainly based on the consideration of prices. Hence, the price development is captured by a variety of indices worldwide, see European Central Bank (ECB), Bank of International Settlements (BIS) and International Monetary Fund (IMF), among others, to assess the state of the residential market. However, not including information about the time it takes to sell or rent out a dwelling might lead to an incorrect assessment of market phases or spatial markets, respectively. To improve the assessment of the residential market, this paper additionally provides quality- and spatial-adjusted liquidity indices for the residential investment and rental market, as complementary demand indices. Hence, this paper aims to provide deeper insights to the under-researched German residential real estate market by combining the consideration of price and liquidity indices. Both indices are developed for the investment and rental market separately for 380 of 401 NUTS 3 regions. The approach is based on the matching of three data sources on an applied big data environment with more than 4.5 million observations, split into about 1.5 million on the investment market and about 3 million on the rental market.

Over the last decade, Germany has experienced a strong and enduring economic expansion. The fundamental economic data exhibits a growing GDP, accompanied by historically high levels of labor demand. The consistently favorable macroeconomic situation and geopolitical events triggered high migration from within the European Union as well as from outside. In addition, the number of households has been increasing due to the social trend towards fewer household members. Furthermore, interest rates for mortgages have been extremely low, resulting in higher affordability of homeownership. Unsurprisingly, this economic and socioeconomic development led to booming demand for residential real estate. Despite rising building permissions and construction activity, building completions have been insufficient to meet demand in many regions. In a study of the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR), Held and Waltersbacher (2015) identified an increased demand of 272 thousand new dwellings per year for the years 2015-2020. The statistics of the Federal Statistical Office (2019a) show that not in a single year since the BBSR published the study, enough new dwellings entered the market (max. 251 thousand in 2018). As consequence, vacancy rates fell below sustainable levels in many regions and house prices as well as rents have experienced upside pressure. The official national house price index of the Federal Statistical Office (2019b) reveals a national price increase of 36% for the last six years. According to the big data sample used for this study, which includes about 1.5 million observations on the investment market,

asking prices increased by 44% on average within the same period.<sup>4</sup> A decomposition of the consumer price index published by the Federal Statistical Office (2019c) reveals an increase in rents of a mere 9.6% for the last six years. Again, the current big data sample of more than 3 million rental offers reveals an increase of 14.9% within the same period.

The price appreciation on the rental market seems innocuous in comparison to the one experienced on the investment market. With a homeownership rate of 43% as of 2013<sup>5</sup>, the first year covered by the current sample, more than half of the German population rent their homes. While Voigtländer (2009), among others, discussed in detail the reasons for the extraordinarily low homeownership rate, research on the nationwide rental market is rather scarce. Furthermore, the ratio of observations on the rental market to observations on the investment market in this paper is about 2:1 and hence demonstrates the importance of the rental market in Germany. Simply based on the moderate appreciation in asking rents, it is hardly possible to make any inferences with regard to a tight rental market. Are the stories about property viewings with more than 50 competitors for the same flat only urban myths? Maybe the analysis of the liquidity on both residential markets reveals the somehow hidden demand. By only looking at the average change in time on market, it seems as there is only a rather small difference between the investment and the rental market, as the liquidity improved by 50% and 39.5% respectively. An estimation of quality- and spatial-adjusted liquidity indices, however, exposes the real difference in market tightness.

However, the price and liquidity development are geographically not equally distributed across the country. A booming trend is mainly observed in the major cities, their surrounding conurbations as well as economically strong regions in southern and north-western parts of Germany. In contrast, rural regions in the east of the country as well as structurally weak regions in the western parts are left behind. Consequently, these areas suffer from the ongoing urbanization and the concomitant rural emigration, leading to low demand and high vacancy rates. Hence, the price and liquidity development differ significantly within Germany. Therefore, it is not sufficient to consider the German market as a whole but to split it up into a regional analysis. That is why, a cluster analysis is conducted in this paper in order to detect economic and socioeconomic similarities and differences of flourishing and fading regions.

The study aims to answer the following questions regarding the residential real estate market.

1. How did price and liquidity on the German residential investment and rental market measured by quality- and spatial-adjusted hedonic indices evolve over the last 6 years?

---

<sup>4</sup> The difference to the Federal Statistical Office might be caused due to the fact that this paper only covers 380 regions in the sample in order to avoid a bias stemming from inactive markets. The Federal Statistical Office, however, includes all regions, and thus the inactive markets that mostly show relatively weak price developments drag the average price development down. Hence, the Federal Statistical Office exhibits a slightly lower price increase compared to our results.

<sup>5</sup> According to GdW (2016)

2. To what extent can the markets be clustered? What similarities and differences in view of economic and socioeconomic indicators do these regions share?

3. Which overall conclusions and implications can be drawn for the German residential real estate market?

The remainder of this paper proceeds as follows: The next section reviews the current state of the literature. Section 2.3 describes the dataset and the descriptive statistics. Section 2.4 presents the econometric model, including the derivation of the hedonic price and liquidity indices as well as the clustering algorithm. Estimation results are presented and discussed in Section 2.5. Section 2.6 concludes.

## Literature

As mentioned above, a deeper understanding of the residential real estate market can only be achieved by examining the price development in combination with the liquidity development. Generally, several liquidity proxies exist for the real estate markets. In line with Sarr and Lybek (2002), Ametefe et al. (2011) summarize various liquidity indicators into five main categories: (1) transaction cost measures, (2) volume-based measures, (3) price impact measures, (4) return-based measures and (5) time-based measures. (1) Transaction cost measures are mainly used for the liquidity analysis on the financial market, see e.g. Clayton and MacKinnon (2000). A market is described as liquid when transaction costs are relatively low. Liquidity is commonly approximated by the spread between the bid and ask price, empirically going back to Demsetz (1968), and related measures. The smaller the spread the more liquid the market. Hence, the bid-ask spread is a measure for market liquidity as well as a transaction cost measure. According to Geltner et al. (2014) the bid price would equal to the asking price, hence, “equilibrium prices” would be charged implying market clearing, if there was a frictionless market with all market participants being perfectly informed about property values. However, when it comes to residential real estate – an asset class which is strongly linked to individual preferences of homebuyers or tenants and expectations of home-sellers or landlords as well as to asymmetric information – both counterparties reveal a different valuation for the same asset. As long as transaction markets are considerably active, both market participants can observe various transaction prices of similar properties. Consequently, they build their beliefs about prices as a distribution around the likely “equilibrium price”. Hence, the higher the transaction frequency on the respective market, the closer the spread between the bid price and the asking price, and in turn the higher the liquidity on the market. To apply this liquidity measure to the German residential real estate market, however, is not possible in this paper due to data restrictions regarding bid prices. (2) Volume-based measures depict liquidity as the absolute or relative number of

transactions. It is mainly used for the liquidity analysis of listed real estate, see e.g. Brounen et al. (2009), however, e.g. Fisher et al. (2009) apply it on the direct real estate market as well. Volume-based measures comprise the transaction volume, the turnover ratio, the quote size, the number of bids and the market depth. The transaction volume is a very popular measure of liquidity in the literature due to its simplicity and data availability. The transaction volume in a specific period is calculated as the sum over all traded assets (price times quantity of an individual asset) in a specific period. Following Geltner et al. (2014), from a theoretical point of view a higher trading volume implies that the assets are generally more similar and, hence, the information of transactions has a greater impact on the appraisal of other assets. Consequently, the distributions of the reservation price of buyers and sellers are converging, decreasing the bid-ask spread as well as the transaction costs and, thus, leading to higher liquidity. Another commonly used liquidity measure is the turnover ratio, however, due to its calculation (transaction volume divided by the product of existing stock and price) is not popular for the direct real estate market as these variables are harder to estimate as e.g. for the stock market. (3) Furthermore, there are different liquidity proxies belonging to the group of price impact measures. These proxies have in common to measure liquidity by separating liquidity from other factors that influence real estate prices. The Amihud (2002) measure for example, applied by Karolyi et al. (2012) as well as Glascock and Lu-Andrews (2013) for instance, is commonly used in the finance literature. Another proxy would be the Pastor-Stambaugh (2003) liquidity factor which is a monthly liquidity measure based on daily data that refers to occasional price changes. (4) Return-based liquidity measures have the advantage that the calculation requires no further information besides the price indices. These measures include for example the illiquidity proxy, developed by Das and Hanouna (2010), which calculates the run length of returns. (5) The liquidity measure used in this paper belongs to the time-based measures. They contain the holding period and the trading frequency, measuring liquidity as the frequency an asset is traded. The holding period, for example applied by Amihud and Mendelson (1986), is empirically applied as the inverse of the turnover rate. For a better suited application in financial markets this liquidity measure was corrected for untraded assets by Collett et al. (2003). Another popular time-based measure is the time on market, measuring liquidity as the time it takes to transact a certain asset. The time on market is especially used in the direct real estate market and is addressed in Krainer (1999) and Allen et al. (2009) among others. The time on market is the measure used in the underlying paper.

As most other markets, the residential real estate market exhibits cyclical movements over time. According to the seminal work of Kluger and Miller (1990) who developed a liquidity measure by using the Cox (1972) proportional hazards model, housing prices and liquidity exhibit a positive correlation. Thus, prices and liquidity should match along “hot” and “cold” market states. Krainer (1999) defines a market as “hot” when prices are increasing, the time on market is short and transaction volume is above average. In contrast, decreasing prices, relatively long selling times and low transaction volumes point to a “cold” housing market. A positive correlation is

found in Berkovec and Goodman Jr. (1996) for instance. Follain and Velz (1995) for example, find a negative correlation. While Stein (1995) and Genesove and Mayer (1997) reason the correlation with sellers' equity constraints, i.e. with frictions on the credit market, Krainer (1999) shows that "hot" and "cold" real estate markets emerge due to search frictions and asymmetric information. Cauley and Pavlov (2002) show evidence for the option value of homeowners and for nominal loss aversion. Substantial deviations from these two market states might indicate speculative expectations by investors and landlords, adjustment processes or supply and demand changes. To detect these deviations is essential for real estate market participants, as it is otherwise impossible to reduce the risk in investment decisions.

Literature in the price-liquidity field focuses predominantly on the US residential investment market. At the same time, academic research concerning real estate market movements on the German market is rather scarce. While most of the literature strand focuses on "hot" and "cold" market phases along the residential cycle, this paper aims to detect "hot" and "cold" market spots on a regional basis. As one of the few papers on the German market, an de Meulen and Mitze (2014) identify "hot" and "cold" spots on the Berlin residential market. In order to detect those, the authors exclusively investigated the price aspect of dwellings. In general, the movements on the residential real estate market are described primarily with price indices. On the overall German market, there are hedonic price indices provided by the Federal Statistical Office as well as indices provided by private companies like e.g. bulwiengesa<sup>6</sup> and Immobilienscout24 (IMX). The methodology and data behind the IMX are described in Bauer et al. (2013). However, a complementary liquidity index and a combination between both are missing. As mentioned above, it is precisely needed to look simultaneously at prices and liquidity when understanding the current and future state of residential real estate markets. Especially for central banks, policy makers, institutional investors, and private households it is essential to be aware of the liquidity momentum, as both indices might move in opposite directions, pointing to different market states. Thus, solely considering the price index for classifying a regional market might lead to incorrect investment strategies and policy implications. Therefore, this paper develops a quality- and spatial-adjusted price and a complementary liquidity indicator for the investment and rental market of 380 German regions. According to the indices, the regional housing markets are then clustered in order to reassess the assumption that prices and liquidity move together or whether their dynamic behavior exhibits frictions. For more than 25 years, bulwiengesa has been providing a clustering of German cities according to their size, measured by the number of inhabitants, the size of the office market and the importance of the city for the national as well as international real estate market. Heinrich and Just (2016) have noted, that those characteristics might not be entirely sufficient for concluding that housing markets form a cluster. While the approach of Heinrich and Just (2016) and the one presented in this paper both use clustering methods, the

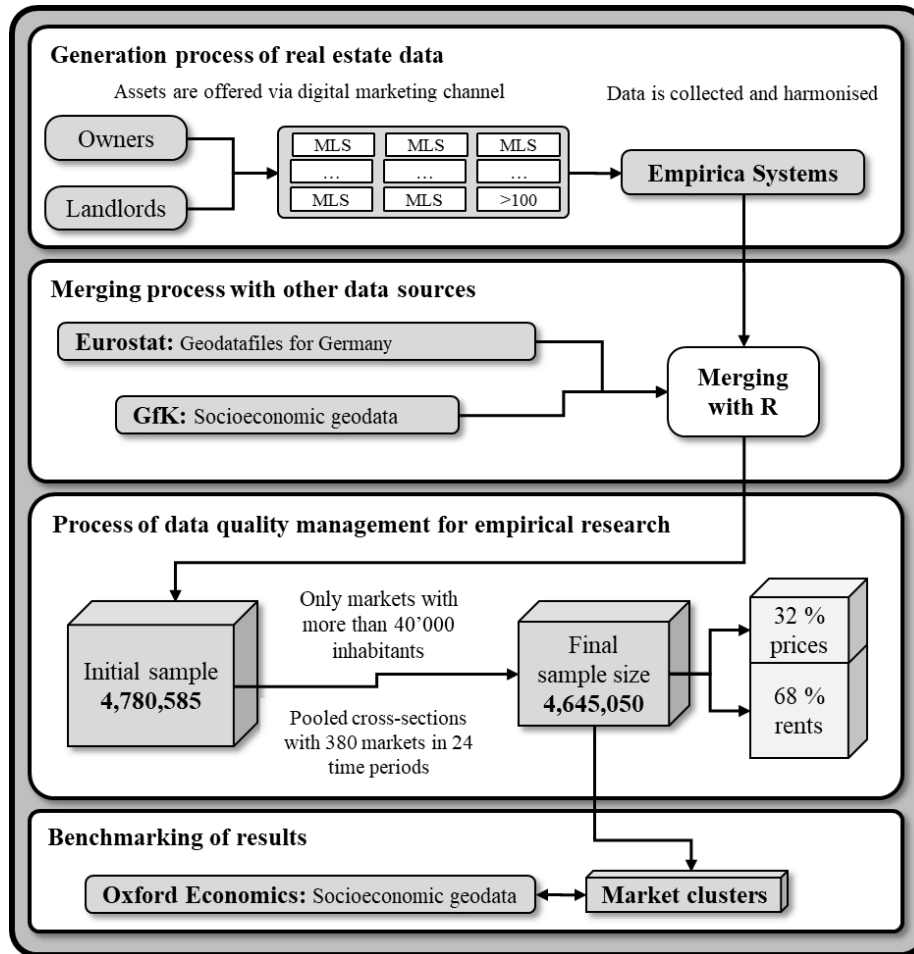
---

<sup>6</sup> bulwiengesa is one of the largest independent consulting companies on the real estate market in continental Europe.

latter one does not directly cluster a variety of variables directly but uses them for the preceding empirical index calculation. In addition to the quality- and spatial-adjusted regional price indices, a liquidity index for each market, respectively, is introduced as an additional clustering indicator. The indexing and clustering on a regional level yields a very granular analysis of the German residential investment and rental market and allows the identification of “hot” and “cold” spots. The findings should be of interest to households, institutional investors and policy makers trying to steer the residential market.

## **Data and descriptive statistics**

In the underlying paper data is gathered from four data sources. Real estate data, including prices on the investment and rental market, are taken from empirica systems ([www.empirica-systeme.de](http://www.empirica-systeme.de)), which collects georeferenced real estate data from more than 100 German Multiple Listing Systems (MLS) such as the market leaders ImmoScout, Immonet or Immowelt but also regionally focused market places and newspapers for the whole German market. As the market leader of real estate data for Germany, empirica has an own proprietary algorithm that identifies doubles and harmonises the sample. Furthermore, the number of households and the purchasing power per household on a ZIP basis are obtained from the “Gesellschaft für Konsumforschung” (GfK) for 2017. The geo-shapefiles of the German territory are extracted from Eurostat in order to calculate two spatial gravity variables: the distance to the centroid of the NUTS 3 region and to the ZIP-centroid. Although the ZIP regions are all part of the larger NUTS 3 regions, the correlation of the variables used for this approach is only 0.311 on the investment market and 0.322 on the rental market. After merging the data and calculating the gravity variables, we proceed with the data selection. This paper only focuses on active residential real estate markets with more than 100 observations and more than 40 thousand inhabitants in order to avoid a bias stemming from a lack of observations on the respective market and outliers. Thus, the number of considered NUTS 3 regions is reduced from 401 to 380 as well as the initial sample size of 4,780,585 observations is reduced to 4,645,050 observations. In total, 1,476,592 observations on the investment market (prices) and 3,168,458 observations on the rental market (rents) in 380 German regions from the first quarter of 2013 to the fourth quarter of 2018 are applied for the estimations. Economic and socioeconomic data for the NUTS 3 regions is gathered from Oxford Economics and is used for the explanation and interpretation of estimation results. The data generating process is depicted in Figure 2.1.

**Figure 2.1** Generation process of residential real estate data

This paper exclusively uses asking prices and asking rents. Due to a lack of transaction prices and contract rents on the German residential market the asking price and asking rent, respectively, operate as a “take it or leave it option” to the buyer or tenant and thus price negotiations are not considered as well. On the residential rental market this assumption is plausible as negotiations about the monthly rental payments are rather an exception. Regarding the property market, Shimizu et al. (2012) compare different house prices – asking prices, contract prices and registry prices – on the Japanese housing market. They conclude that for the different house prices and, thus, different datasets there exist differences in the housing price distribution, which can be traced back to quality differences. However, after controlling for quality differences across the datasets no substantial differences between the house price distributions are to observe. Another peculiarity of this study is the application of the total price in Euro of a dwelling offered on the investment market and the total monthly rent in Euro for a dwelling offered on the rental market, although the more popular approach is the utilization of price and rent per square meter. When estimating time on market, we observed that the application of those commonly used variables yields economically incomprehensible relationships for the impact of changes in price (improves liquidity all else equal) and living area (decreases liquidity all else equal) on liquidity. The time

on market, is defined as a non-negative continuous variable, measuring the time elapse that a dwelling requires to change its status from being offered on the market to being out of the market in weeks calculated by its start and end date. Typical housing attributes included in this study are hedonic characteristics like “living area”, “age” and “number of rooms” as well as binary hedonic characteristics like for example “with balcony”, “with parking slot”, “with elevator”. Since the data is georeferenced, NUTS 3 regions have been employed as geographical analysis units and are defined by Eurostat to the “Nomenclature of Territorial Units for Statistics”, which is a hierarchical system for dividing the economic territory in Europe. Hence, this classification provides the possibility of statistical comparison of regions within the EU. NUTS 3 regions are the smallest classification units, thus cover small regions that are related to counties or administrative districts. This classification of spatial territory is strongly related to the administrative division of the country. Subsidies for specific regions are also assigned according to the classification of NUTS regions. Michels et al. (2011) criticize that this type of classification exhibits the administrative structures, however, cannot depict economic and functional linkages to the surrounding areas. They propose to classify the regional markets into “housing market areas”, defined as areas where a household lives and works and within that area the household will search for an alternative place to live in the case of relocation. Hence, Michels et al. (2011) take migration as well as commuting flows into account and make their classification more functional. Even though the classification into “housing market areas” constitutes a very effective approach, it is not applicable to the underlying paper. The aim of this paper is to allocate regions to clusters according to their price and liquidity development and further to detect economic and socioeconomic similarities of these regions. Consequently, an official classification is inevitable in order to match the clustered regions to economic and socioeconomic data based on the same regional level. Furthermore, the classification of “housing market areas” might lead to a different classification on the investment and regional market. Hence, this could lead to further complications regarding the underlying paper.

The variables, their units and sources can be found in Table 2.1.

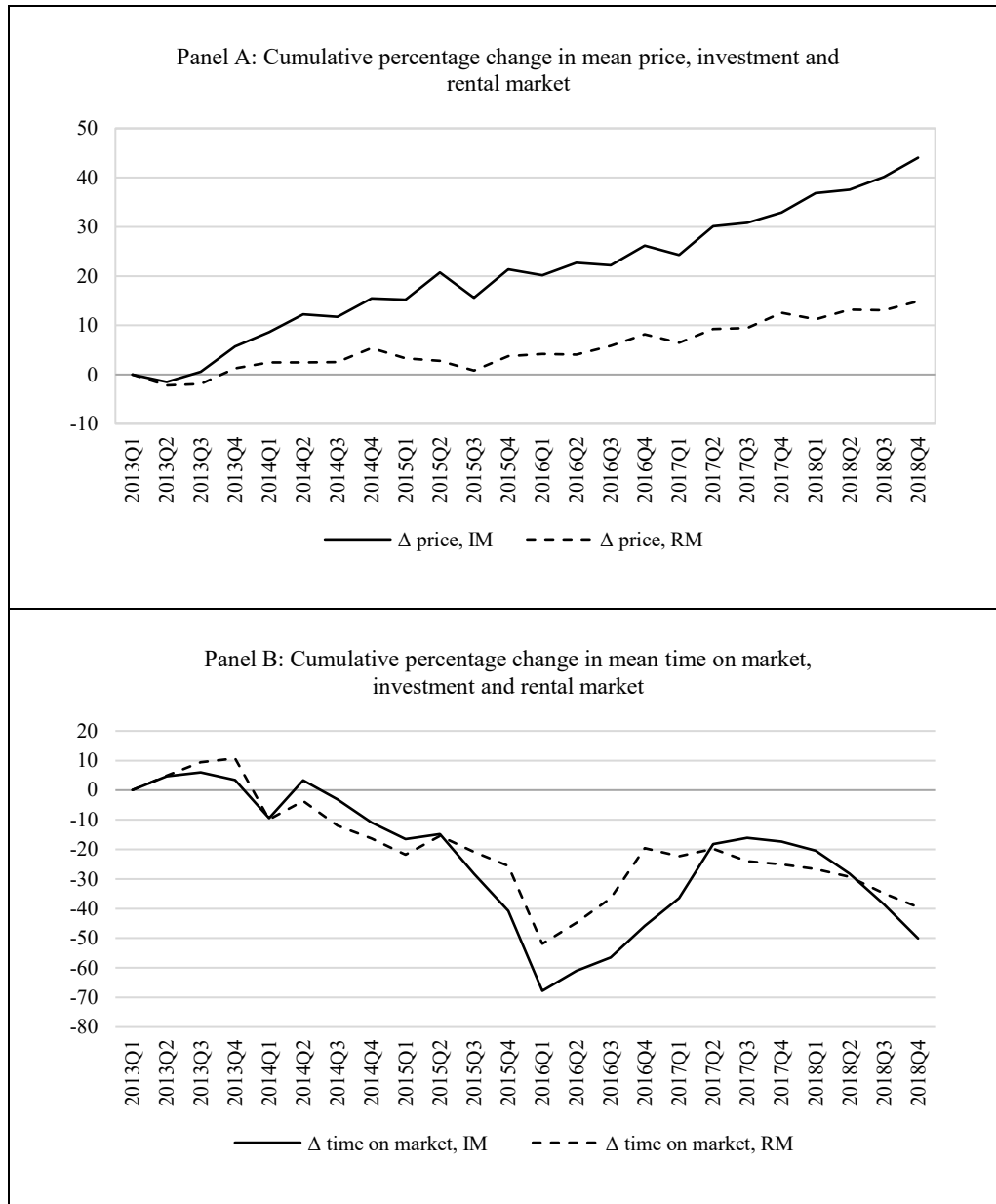
**Table 2.1** Variables and sources

Variable	Unit	Effect in the estimation equations			Sources		
		Hedonic effects	Spatial effects	Socioeconomic effects	Empirica	GfK	Eurostat
Asking price	€	✓					
Asking rent	€ per month	✓			✓		
Time on market	weeks	✓			✓		
Living area	m <sup>2</sup>	✓			✓		
Age	years	✓			✓		
Rooms	number	✓			✓		
With bathtub	binary	✓			✓		
With built-in kitchen	binary	✓			✓		
With car space	binary	✓			✓		
With terrace	binary	✓			✓		
With balcony	binary	✓			✓		
With elevator	binary	✓			✓		
Newly built dwelling	binary	✓			✓		
Refurbished dwelling	binary	✓			✓		
Gaussian longitude	coordinate		✓		✓		
Gaussian latitude	coordinate		✓		✓		
Distance to ZIP centroid	km		✓				✓
Distance to NUTS 3 centroid	km		✓				✓
Households in ZIP	HHs/ZIP			✓		✓	
Purchasing power of HHs in ZIP	€/HH/p.a./ZIP			✓		✓	
N							4,645,050

Figure 2.2 shows the cumulative mean price and time on market development on the investment (IM) and rental market (RM) from the first quarter of 2013 until the end of the observation period. It is visible that prices have been increasing accompanied by a diminishing time on market on

both markets. Hence, both indicators point to a boom phase on the German housing market, triggered by ongoing demand with supply lagging behind. Moreover, it is observable that prices on the transaction market have been increasing considerably more than rents. While rents have been rising by a mere 14.9%, prices on the transaction market have experienced a substantial growth of 44% over the last six years. Those figures indicate a particularly high demand on the investment market, probably triggered by constantly low mortgage rates on housing loans and a lack of interest-bearing investment opportunities.

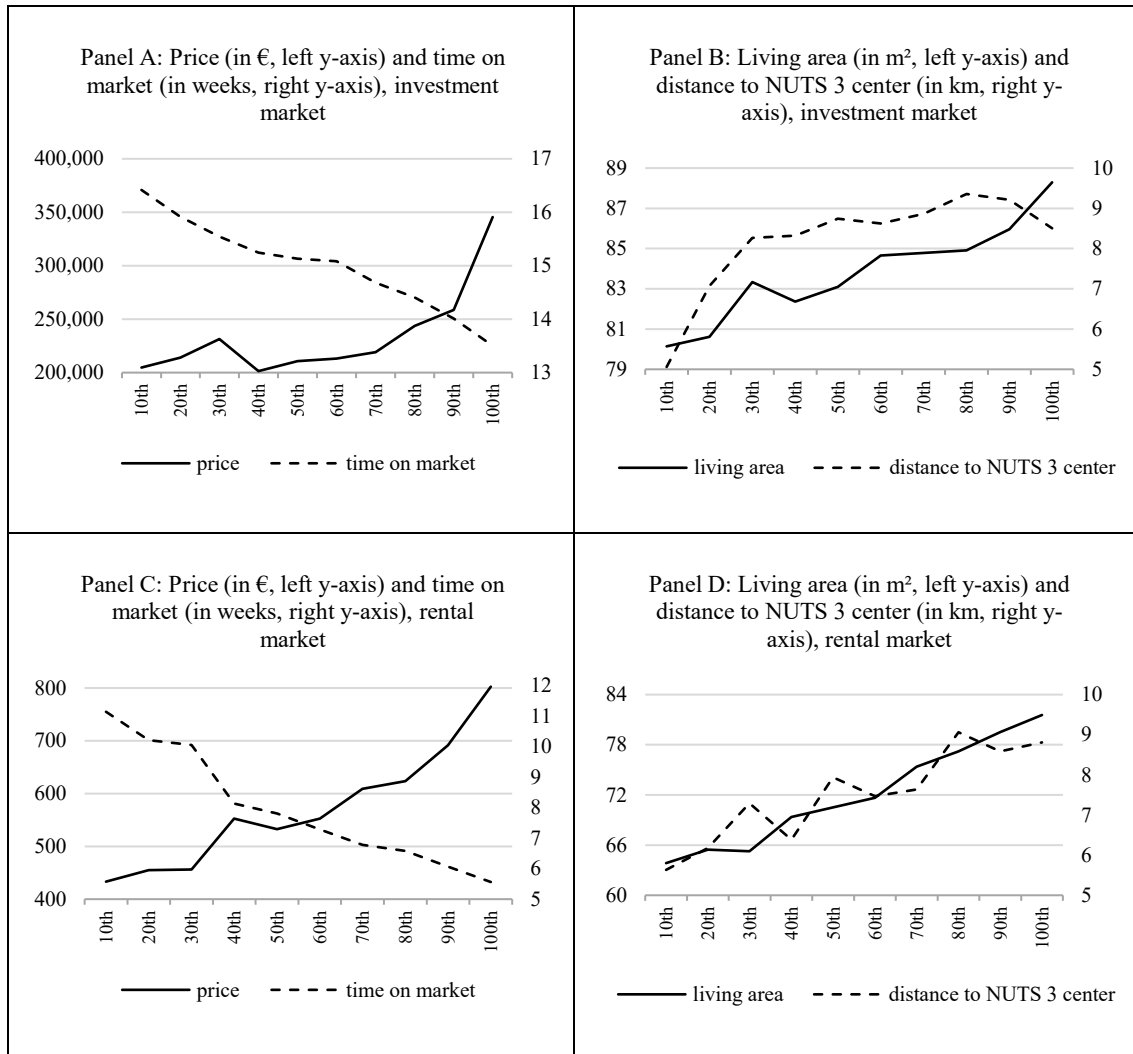
**Figure 2.2** Cumulative percentage change in mean price and mean time on market



**Notes:** This figure plots the cumulative percentage change in mean price on the residential investment and rental market as well as the cumulative percentage change in time on market on the residential investment and rental market. The data consists of 1,476,592 observations on the residential investment market and 3,168,458 observations on the residential rental market. The sample period is 2013 Q1 to 2018 Q4.

It seems that the price development has not yet been fully absorbed by the rental market. The relatively moderate growth in rents seems to only reflect the natural demand, which obviously was higher in cities. As landlords will try to pass on the rising prices on the investment market to their tenants, it might indicate further rental growth in the near future. Of course, rental protection laws prohibit landlords to hand over the entire increase in transaction prices to tenants in order to meet their target return. Asking exorbitant rents has been prohibited on the German market for years, not only since the introduction of rent control in 2015. Because of lacking investment alternatives, new landlords somehow became acquainted to shrinking rental yields. Nevertheless, time on market exhibits an enormous decrease of about 50% on the transaction market and about 40% on the rental market with an almost parallel development on both markets. Although prices on the investment and rental market have not experienced growth of equal magnitude, the similar development of time on market indicates considerably high demand on the rental market, which might also result in upward pressure on rents. To reason the similar drop in time on market with relatively more supply on the rental market in relation to the investment market is rather less plausible, as newly built dwellings are usually offered on the investment market, before they appear on the rental market. Thus, this slightly controversial finding emphasizes the importance of focusing on both indicators – price and time on market – when analysing the residential real estate market.

Figure 2.3 exhibits, that heterogeneity is omnipresent on the housing market. Panel A to D show, that households within different purchasing power percentiles demand a different price, living area, and distance to the city center. Furthermore, it is shown that the sales and letting process with respect to the marketing time varies.

**Figure 2.3** Distribution of selected variables across purchasing power percentiles

**Notes:** The figures plot the distribution of selected variables segmented by purchasing power percentiles. The data consists of 1,476,592 observations on the residential investment market and 3,168,458 observations on the residential rental market. The sample period is 2013 Q1 to 2018 Q4.

Generally, relatively richer households can afford more expensive dwellings, prefer larger living areas, tend to live further away from the city center and spend less time on the search and matching process. These preferences are visible for the investment as well as the rental market. Surprisingly, buyers living within zip codes with very low purchasing power (20<sup>th</sup>-, 30<sup>th</sup> percentile) pay on average higher prices than the middle-income (40<sup>th</sup>-, 50<sup>th</sup>-, 60<sup>th</sup> percentile) groups. On the investment market a reason for that might be that the living area demanded by households at the 40<sup>th</sup> and 50<sup>th</sup> purchasing power percentile is lower than at the 30<sup>th</sup> percentile and furthermore, these households live further away from the city center. However, this cannot be observed on the rental market. Another interesting fact is, that the range between the highest and lowest income group with respect to prices, living area and time on market is remarkably more pronounced on the rental market relative to the investment market. While on the investment market asking prices, living area and the time on market between the richest and poorest percentile vary by 68.7%,

10.2% and 17.8%, the differences on the rental market are much stronger with 85.3%, 27.7% and 50.1% respectively. This infers that the participants and probably also the dwellings on the rental market are much more diversified than those on the investment market. It seems surprising, that relatively rich households tend to spend much more on renting than on the investment in a dwelling compared to relatively poor households. Regarding the distance to the city center it is noticeable that especially among the poorest, an increase in purchasing power leads to a very strong shift of investments further away from the city center.

## Econometric approach

The aim of a price index is to measure the price development over successive periods after correcting for hedonic characteristics. However, residential dwellings are not transacted periodically, but rather irregularly and even infrequently. Furthermore, residential real estate is extremely heterogeneous, both in terms of its physical characteristics and its location. Dwellings with different characteristics and in different locations might exhibit distinct price and liquidity dynamics in terms of volatility and cyclicity. Thus, idiosyncratic price and liquidity movements might be to observe in diverse markets, due to social, and economic circumstances in a particular region. In order to control for heterogeneity, hedonic indexing is applied in this paper. The hedonic approach is a method for generally indexing economic prices of goods affected by quality changes. Kain and Quigley (1970) were among the first to apply hedonic pricing to the real estate market. Given hedonic data, the hedonic model decomposes the price as well as the liquidity of residential real estate into individual characteristics. Hence, the computed index reveals constant characteristics and consequently points out the pure price and liquidity changes over time. The location of a dwelling is probably one of the most important determinants of prices and liquidity. Therefore, not only postcode identifiers as well as longitude and latitude data are considered in the functional form, but the price and liquidity indices are estimated individually for each market  $p \in \{1, \dots, 380\}$ , defined by the NUTS 3 regions. In this paper, the time-dummy method is applied, which is defined as the marginal change in price (liquidity) with respect to time. Thus, a transformation of the estimated coefficients of the time fixed effects yields the price (liquidity) index, referring to the percentage marginal change in prices (liquidity) in period  $t_t$  relative to period  $t_0$ . Hence, the indices can be computed directly from the estimated coefficients. Compared to the imputed hedonic index no “representative dwelling” must be defined and it is less data intensive and therefore very well suited for the construction of regional price and liquidity indices. The standard model for the estimation of a time-dummy hedonic index is given as

$$y = X\beta + \mu\theta + u.$$

(1)

As the semi-log functional form has proven appropriate and is used in most hedonic regression models according to Halvorsen and Pollakowski (1981) as well as Diewert (2003), among others,  $y$  is an  $i$ -vector consisting of the elements  $y_i = \ln(p_i)$ .  $i$  denotes the number of dwellings in the sample.  $X$  is defined as an  $(i \times C)$ -matrix of covariates, with  $C$  being the number of covariates without the time dummies,  $\beta$  is a  $C$ -vector, describing the shadow price of each covariate. To generate an intercept as the first item of  $\beta$ , the first column of  $X$  solely consists of ones.  $\mu$  is an  $(i \times T-1)$ -matrix of time dummies for each period, with  $T$  being the number of observation periods,  $\theta$  is a  $(T-1)$ -vector of period shadow prices relative to a fixed time period  $t_0$ , and  $u$  is an  $i$ -vector of error terms. As the purpose is to generate a price index, the coefficient of interest is the time dummy parameter  $\theta$ .  $\theta$  quantifies the time period-specific fixed effects, i.e. the impact of each time period, on the log price after controlling for quality and spatial characteristics of a dwelling. Exponentiating the estimated coefficient  $\hat{\theta}_t$ , yields the time-dummy index as

$$\widehat{P}_t = \exp(\hat{\theta}_t). \quad (2)$$

A transformation via  $[\exp(\hat{\theta}_t) - 1] \cdot 100$  corresponds to the marginal change in prices in  $t_t$  relative to  $t_0$ . It is to note, that the time dummy index estimated above is not unbiased. According to Goldberger (1968) for example,

$$\widehat{P}_t^* = \exp\left(\hat{\theta}_t + 0.5 \left(\widehat{se}(\hat{\theta}_t)\right)^2\right), \quad (3)$$

yields a standard bias correction.  $se(\theta_t)$  refers to the standard error of the time-dummy coefficient. However, according to Goldberger (1968) and Syed et al. (2008), among others, the bias is in general very irrelevant with Syed et al. (2008) showing that the difference in the indices appears only in the fourth decimal place. Thus, there is no need to correct for the bias according to Triplett (2004) and de Haan (2010), among others.

As this paper aims to investigate the dynamics of prices and liquidities, four models are estimated in order to obtain the price index on the investment as well as rental market and the two liquidity indices for the investment and the rental market. While for the price indices hedonic regressions are estimated, survival models are set up to obtain the liquidity indices. The four models are estimated individually for each NUTS 3 market  $p \in \{1, \dots, 380\}$  as independently pooled cross-sectional regressions.

### 2.1.1 The residential price index

This section describes the derivation of the time-dummy price index for the residential real estate investment as well as rental market. The hedonic equation (4) is estimated for the investment and rental markets separately based on the approach of Cajias (2018). Estimation is conducted via a semiparametric Generalized Additive Model for Location, Scale and Shape (GAMLSS)

introduced by Rigby and Stasinopoulos (2005). The main reason for the usage of the GAMLSS approach is the fact that prices on the real estate investment and rental market vary across space, time and within submarkets. The approach models the parameters of the response as semiparametric functions of the covariates and expands the regression equation by considering the four moments of the response – the mean, variance, skewness and the kurtosis – in the optimization algorithm. The GAMLSS approach is widely recognized and used by international institutions such as the International Monetary Fund, the World Health Organization or the European Commission. The models are parameterized for the price as follows:

$$\ln(P_{ijt}) = X_i\beta + Z_{jt}\alpha + \mu_t\theta_t + u_{ijt}. \quad (4)$$

The hedonic regression decomposes the log price  $P$  of a dwelling  $i$  in ZIP-code  $j$  and in observation period  $t$  into dwelling-specific characteristics  $X_i$  and ZIP-code-specific covariates  $Z_{jt}$ .<sup>7</sup>  $\mu_t$  captures the time fixed effects, thus is the focus of the index calculation. The error term  $u_{ijt}$  describes the variation in prices that cannot be explained by the model. In this case independently and identically distributed ( $u \sim iid$ ) robust standard errors are used for the regression. As the time dummy index is defined as the marginal change in price  $P_{ijt}$  with respect to  $\mu_t$ , a transformation of the estimated coefficients  $\hat{\theta}_t$  according to

$$\widehat{PI}_t = [\exp(\hat{\theta}_t) - 1] * 100, \quad (5)$$

yields the price index  $PI_t$ , referring to the percentage marginal change in prices in period  $t_t$  relative to  $t_0$ .

### 2.1.2 The residential liquidity index

Without any doubt the leading model for the analysis of survival data is the Cox (1972) Proportional Hazards Model (PHM). This model is used for exploring the determinants of the duration of an event or elapse of time, e.g. it determines the variables that accelerate or restrict the elapse of time that a response variable needs to change its state. In this case, the response variable is defined as a non-negative continuous variable, measuring the elapse of time that a dwelling requires for changing its status from being offered on the market into being out of the market in weeks, i.e. time on market. For understanding and estimating survival data, two main functions are essential: the survival function  $S(t)$  and the hazard rate function  $\lambda(t)$ . The survival function specifies the probability that an event has not occurred until a certain time  $t$  and is formally defined as

---

<sup>7</sup> Since the correlation of the variables “living area” and “number of rooms” is at about 0.6, we decided to analyze whether the inclusion of both variables improves the explanatory power of the regressions estimating the price and liquidity indices. The results show that the explanatory power decreases by 8 to 14% for the different indices if only “living area” is included.

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^{\infty} f(x)dx, \quad (6)$$

with  $f(x)$  being the probability density function of the time until the event. The hazard function  $\lambda(t)$ , in contrast, describes the probability at  $t$  that an event occurs at time  $T$ , given that the event has not occurred before and is given by

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | t \leq T)}{\Delta t}. \quad (7)$$

The relationship between those two functions is straightforward since the integrated hazard rate  $\Lambda(t) = \int_0^t \lambda(x)dx$  can be expressed as the negative log of the survival rate  $S(t)$  as  $\Lambda(t) = -\log S(t)$ . In other words, the survival function expresses the probability of a dwelling for staying in the market while the hazard function measures the risk of the same dwelling for leaving the market.

The Cox PHM estimates the survival function, but coefficients can be transformed to hazard rates, giving the probability of “mortality” per unit of time, and hence describing a liquidity indicator. The semiparametric Cox proportional hazards regression is parameterized as

$$\lambda(\tilde{t}_{ijt}) = \tilde{X}_i \tilde{\beta} + Z_{jt} \tilde{\alpha} + \mu_t \tilde{\theta}_t + e_{ijt}. \quad (8)$$

The time on market  $\tilde{t}$  of a dwelling  $i$  in ZIP-code  $j$  and in observation period  $t$  is decomposed into dwelling-specific characteristics  $\tilde{X}_i$  and ZIP-code-specific covariates  $Z_{jt}$ . In addition to  $X$ ,  $\tilde{X}$  includes the log of asking prices as the data generating process (DGP) of the time on market  $\tilde{t}$  is influenced by the initial asking price, as landlords set the asking price when offering the dwelling in the MLS. As in the hedonic survival regression, time fixed effects  $\mu_t$  are included and  $e_{ijt}$  describes the error term. With  $\exp(\tilde{\theta}_t)$  being defined as the hazard rate, the estimated coefficients  $\tilde{\theta}_t$  can be transformed into the liquidity index  $LI_t$  according to

$$\hat{LI}_t = [\exp(\tilde{\theta}_t) - 1] * 100, \quad (9)$$

referring to the percentage marginal change in the hazard rate, i.e. in liquidity, in period  $t_t$  relative to  $t_0$ .

### 2.1.3 Cluster analysis

In order to determine regional markets that coincide according to their market movements, proceeding from the price and liquidity indices, the 380 regions are assigned to one of two clusters. The clustering is conducted separately for the price and liquidity indices on the transaction and rental market. Hence, the clustering is conducted on four sub-datasets. The aim of the cluster analysis is to assign regions to the same cluster, so that the dissimilarity within a cluster is minimized and maximized between the clusters. Therefore, the “partitioning around medoids (PAM)” clustering algorithm, going back to Kaufman and Rousseeuw (1987), is applied. The PAM clustering algorithm belongs to the  $k$ -medoids clustering procedure, that is closely

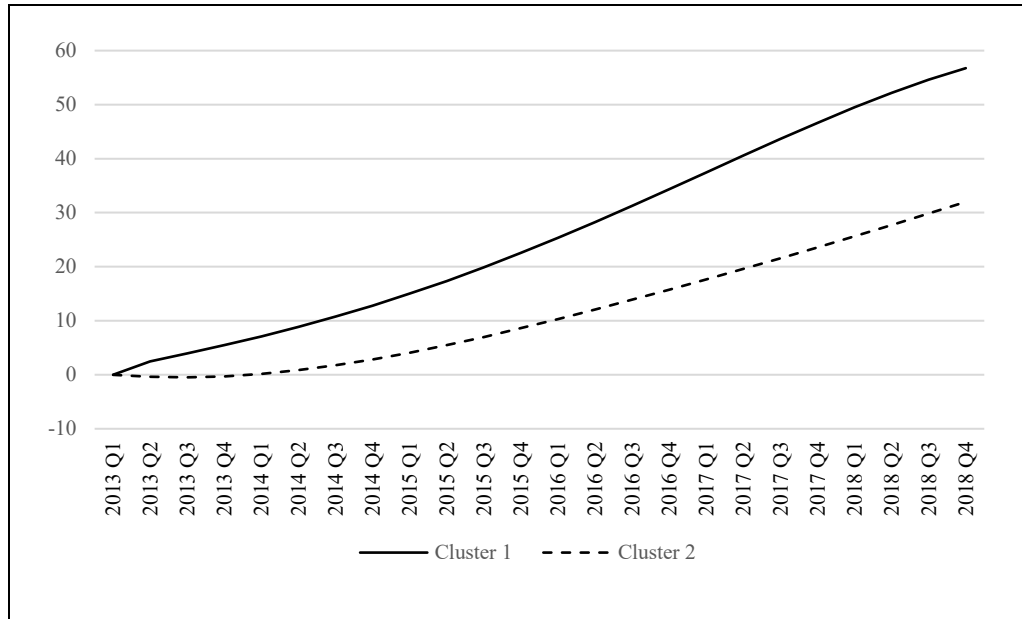
related to the  $k$ -means procedure, however, according to Kaufman and Rousseeuw (1990), is more robust to outliers and noise. While the  $k$ -means algorithm aims to minimize the sum of squared Euclidean distances, the  $k$ -medoids algorithm minimizes the average dissimilarity between the “representative” object, i.e. the medoid, and all other objects of the respective cluster. As with all partitioning methods, the PAM clustering algorithm requires to specify the number of clusters  $k$  a priori. Similar to Just et al. (2019) at first the optimal number of clusters is identified. According to Kaufman and Rousseeuw (1990), in this study the average silhouette method, providing an evaluation of the quality of a clustering, is applied. It identifies how well each observation fits into a cluster. This approach computes the average silhouette of observations for several different numbers of clusters. The number of clusters where the average silhouette width is maximized is optimal. For each of the four sub-datasets used in this paper the optimal number of clusters is two. The PAM algorithm consists of two major steps, the BUILD phase and the SWAP phase. At first,  $k$  initial objects are selected as medoids, i.e. these objects minimize the sum of the distances to all other objects. Second, the objective is to optimize the set of medoids. Therefore, each pair of medoid and remaining object is exchanged. If a swap indeed improves the cluster quality, the initial medoid and the other object change positions. This iteration is conducted until the quality of each cluster is optimal. The stability of the cluster solution provided by the PAM algorithm is assessed via bootstrap. The clusterwise Jaccard similarities suggest that each cluster is stable. The decisive variables underlying the clustering procedure are the estimated time-dummy coefficients  $\theta_t$  and  $\tilde{\theta}_t$  at each observation period  $t$ . Based on this clustering analysis it is the aim of this paper to identify “hot” and “cold” regions as well as regions expected to ascend and descend in terms of price development.

## Regional cluster results

### 2.1.4 Cluster results on the investment market

After estimating the price and liquidity indices for the 380 NUTS 3 regions, each region is assigned to one of two clusters according to the methodology described in section 2.4.3. Berlin for example, is assigned to cluster 1 for its price development and cluster 2 for its liquidity development. In the following, the city will be referred to as Berlin (1,2)<sup>8</sup>.

**Figure 2.4** Price index, investment market



**Notes:** The figure plots the mean cumulative percentage price change for dwellings allocated to the individual price clusters. The price changes are presented as the coefficients of the time dummy variable of a quality- and spatial-adjusted GAMLSS regression. To cluster the index values, the partitioning around medoids (PAM) algorithm was used. The data consists of 1,476,592 observations on the residential investment market. The sample period is 2013 Q1 to 2018 Q4.

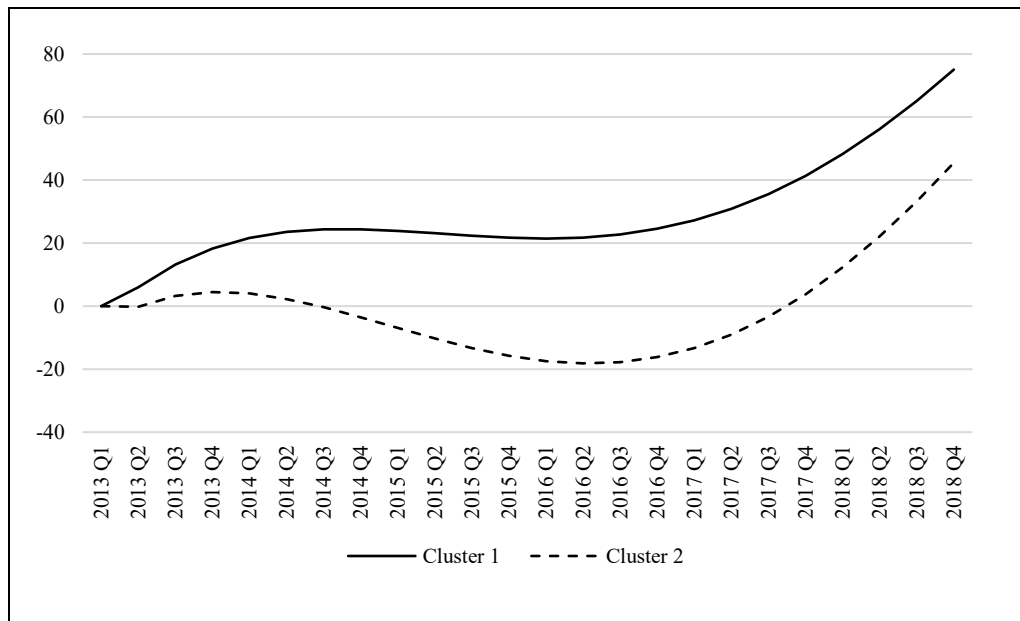
The trend of the quality- and spatial-adjusted price cluster means is shown in Figure 2.4 and reveals the cumulated average price change of all dwellings allocated to the specific price clusters, indexed to zero in 2013 Q1. While NUTS 3 regions allocated to price cluster 2 experienced slightly decreasing prices compared to the base quarter at the beginning and started to increase from 2014 Q1 onwards, for price cluster 1 a consistently positive price development is visible. However, over the entire observation period both price clusters display a quite similar upward-sloping trend. Over the past six years, prices in price cluster 2 have been rising on average by 32% and in price cluster 1 even by 56.75%. As the price development in price cluster 1 is much steeper, these regions can be identified as highly demanded regions relative to regions allocated to price cluster 2.

<sup>8</sup> A full list containing the classification of the 380 NUTS 3 regions is available upon request.

Comparing regions assigned to price cluster 1 versus regions assigned to price cluster 2 by means of economic and socioeconomic data shows, that at the median<sup>9</sup> regions in price cluster 1 are larger with respect to the population as well as the working population. An even more pronounced trigger of the very strong price development in price cluster 1 might be the positive development of the population and working population over the whole observation period. Price cluster 1 regions experienced a 3.84% increase in population and working population rose by 2.5%. This definitely implies a higher demand for living space within the respective regions, resulting in higher prices. In contrast, price cluster 2 regions only had a very small increase in population of 0.97% and even suffered a loss of 0.91% in working population. This very small increase in demand might to some extent be responsible for the lower price development in price cluster 2. However, as prices in price cluster 2 have also been rising not negligibly, other price drivers than the pure demand for space must be an issue. Regions allocated to price cluster 1 are furthermore characterized by a higher real GDP, lower unemployment rates and a higher disposable household income compared to price cluster 2. While in 2018 the median real GDP in price cluster 1 was about €5.2 million, regions in price cluster 2 only had a median real GDP of €4.3 million. It is not surprising, that prices have had a larger growth in more productive regions as this factor is a strong demand indicator and thus correlates with population changes. From 2013 to 2018 real GDP growth is observed for both price clusters, with the median real GDP in price cluster 1 increasing by 10.3% and 8.42% in price cluster 2. The generally flourishing economic conditions can also be observed by decreasing unemployment rates and increasing disposable household income. While the unemployment rate of 2.67% in price cluster 1 is lower than in price cluster 2 with 3.28%, it is to note that the decline in the unemployment rate is more pronounced in price cluster 2. The same is to observe for the disposable household income. While the median disposable household income of €48.43 thousand in price cluster 1 is higher than in price cluster 2 with €43.29 thousand, the increase in price cluster 2 is stronger. The relatively stronger development of the unemployment rate and the disposable household income in price cluster 2 is not surprising, as these regions started with relatively weak levels for the calculation of the growth rates. The relative strong decrease in the unemployment rate in price cluster 2 and the relative strong increase in disposable household income in combination with real GDP growth might somehow indicate that smaller regions, in terms of population, with still very low disposable household income are in the progress of slowly adjusting to the larger and economically stronger regions. This favourable economic development in rather small regions contributed to the upward sloping price development in price cluster 2 without strong gains in population and working population. There is no doubt that the prevalent situation on the investment market in general, with very low mortgage rates and sparse alternative investment opportunities played a significant role for the overall positive price development.

---

<sup>9</sup> The median is used instead of the mean in order to prevent outliers from distorting the cluster values.

**Figure 2.5** Liquidity index, investment market

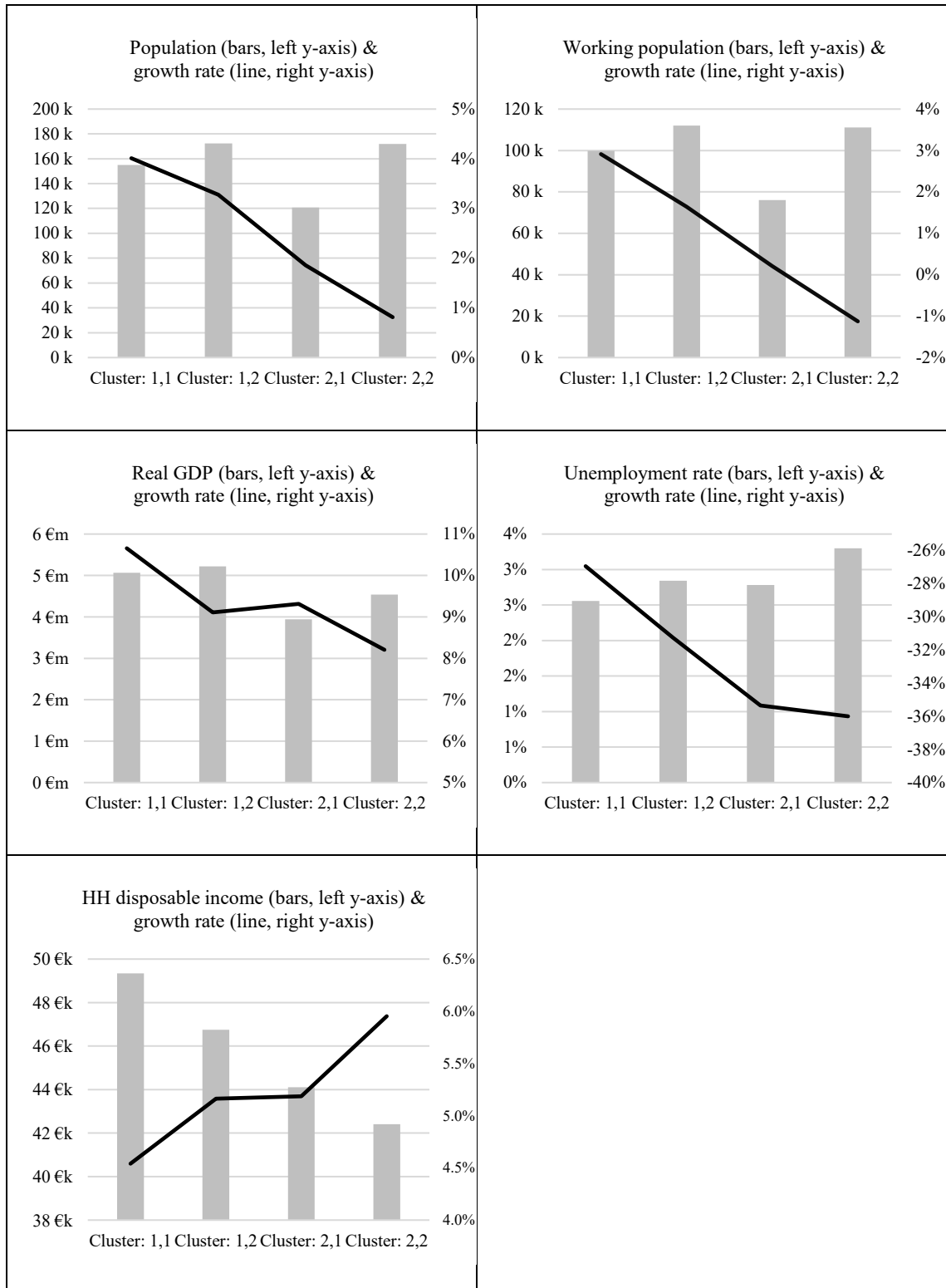
**Notes:** The figure plots the mean cumulative percentage change in liquidity for dwellings allocated to the individual liquidity clusters. The changes are presented as the coefficients of the time dummy variable of a quality- and spatial-adjusted Cox proportional hazards model. To cluster the index values, the partitioning around medoids (PAM) algorithm was used. The data consists of 1,476,592 observations on the residential investment market. The sample period is 2013 Q1 to 2018 Q4.

Assigning the regions to clusters by their liquidity development, which is based on the time it takes to sell a dwelling within the respective regions, displays a different pattern but also higher index values at the end of the observation period. As shown in Figure 2.5, for regions assigned to liquidity cluster 2, liquidity was worse than in the base quarter in 14 out of 23 quarters. From 2014 Q1 onwards liquidity in liquidity cluster 2 was declining, i.e. marketing time in the respective regions was getting longer. Liquidity hit the bottom in 2016 Q2, exceeded the level of the base quarter in 2017 Q4 and finished with a plus of 45.48% in 2018 Q4. Regions assigned to liquidity cluster 1 experienced a similar process. Though liquidity in liquidity cluster 1 never declined below the level of the base quarter, it was rather stagnating from 2014 Q1 until the end of 2016. The stagnating liquidity development might to some extent be caused by the sharp increase in building completions between 2013 and 2016. According to the Federal Statistical Office (2019a) building completion of dwellings has been rising from about 188 thousand in 2013 to about 236 thousand in 2016. As in this period the high demand for dwellings faces increased supply, liquidity remains rather constant. Afterwards liquidity experienced a sharp increase, finishing with a plus of 75.05% in 2018 Q4. Hence, over the entire observation period dwellings became on average more liquid in both liquidity clusters.

Taking a look at the economic and socioeconomic data of the considered regions, it is to observe, contrary to the price clustering, that liquidity was increasing most in regions with a lower median population as well as working population. The difference in population between liquidity cluster 1 and cluster 2 is way less pronounced than between the price clusters. The growth rates of these

two variables, however, support the relative stronger liquidity growth in liquidity cluster 1. While the population in liquidity cluster 1 rose by 3.39% and the working population by 2.14%, liquidity cluster 2 experienced a rather small increase of 1.66% and 0.1%, respectively. As the plus in population generates more demand for space, the housing market gets tighter, hence the time it takes to sell a dwelling has been decreasing. In line with population, regions allocated to liquidity cluster 1 show a slightly smaller real GDP of €4.66 million compared to liquidity cluster 2 with €4.76 million. In comparison with price clustering, liquidity clustering results in way more similar clusters with respect to real GDP. The change in real GDP, that is more decisive when it comes to liquidity development, is higher in liquidity cluster 1 with 10.5% compared to liquidity cluster 2 with 8.79%. As with the price clustering these productivity gains coincide with the population development and hence result in higher demand for space leading to higher liquidity levels. For the unemployment rate, household disposable income as well as the respective growth rates the same picture as with price clustering is to observe. The increase in liquidity was stronger in regions with a lower unemployment rate and a higher disposable income. However, what might have triggered the liquidity growth in liquidity cluster 2, though new demand for space in terms of population gains was relatively weak, was the pleasant development of the unemployment rate and disposable household income.

**Figure 2.6** Socioeconomic data by cluster affiliation, investment market



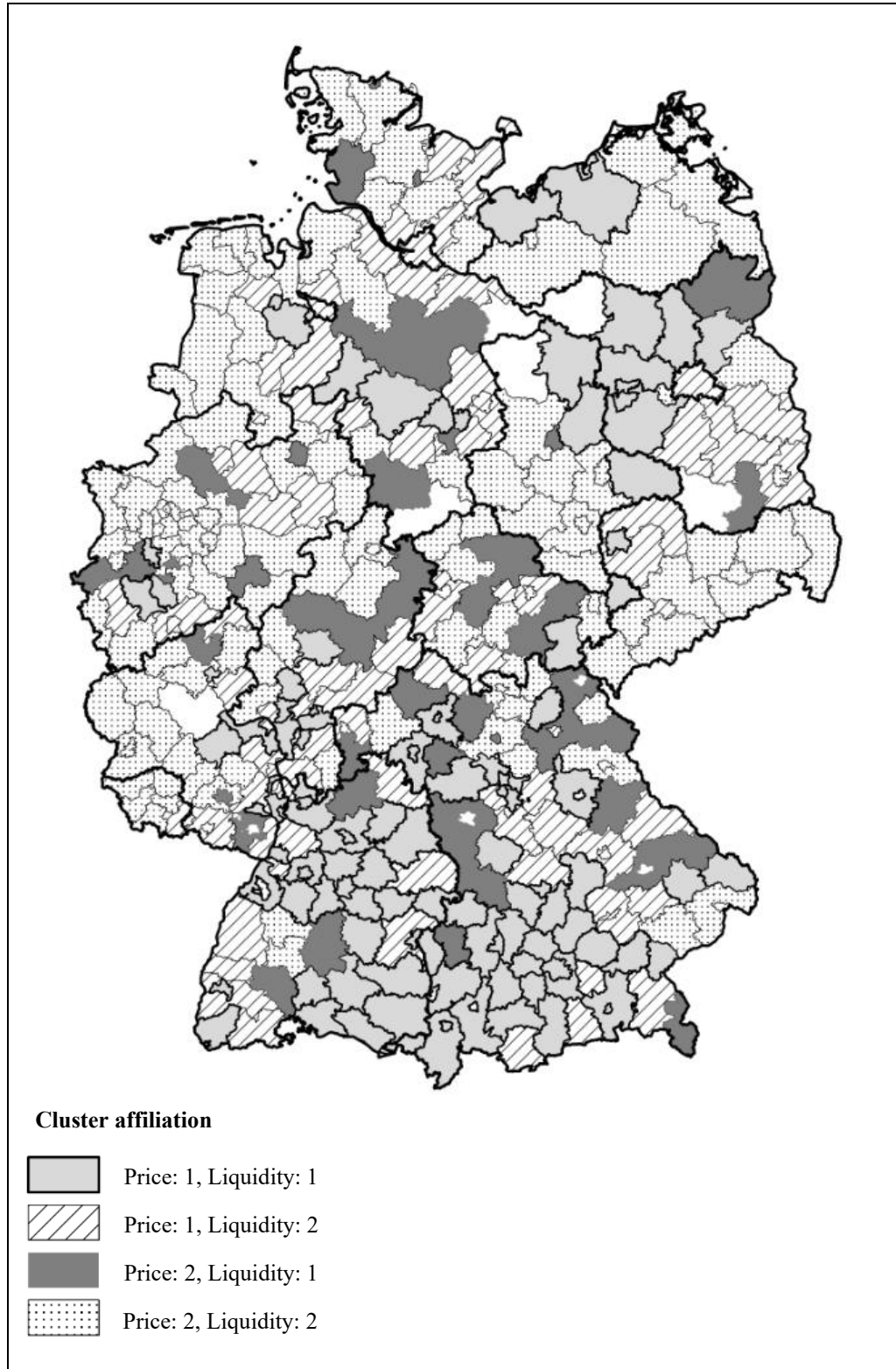
**Notes:** The figure displays the socioeconomic data by cluster on the investment market. The NUTS 3 regions are allocated to a specific cluster by applying the partitioning around medoids (PAM) algorithm on the price and liquidity index values, respectively. The grey bars exhibit the actual values of the respective variables, labelled on the left y-axis. The black lines display the respective growth rates and are labelled on the right y-axis. The data consists of 1,476,592 observations on the residential investment market. The sample period is 2013 Q1 to 2018 Q4.

As mentioned above, a combination of the cluster ranks derived from the price and liquidity development is used in order to classify the 380 regions. For this purpose, the price cluster rank

is regarded as the primary determinant and the liquidity cluster rank as the complementary secondary determinant, which enables a higher granularity in the classification and a more precise market assessment. Out of the 199 regions allocated to cluster 1 by price, 109 are as well allocated to liquidity cluster 1. Those regions can be declared as absolute “hot” markets, where an extraordinary price development is supported by a very strong liquidity development. The 128 regions assigned to cluster (2,2) are characterized as “cold”. Cluster (1,2) as well as cluster (2,1) can be found in the middle. Figure 2.6 summarizes the economic and socioeconomic data for the four possible cluster combinations on the investment market.

It is to observe that population and working population are the highest in regions with a relative weaker liquidity development irrespective of the price cluster, i.e. clusters (1,2) and (2,2). These two clusters, however, also exhibit the highest unemployment rates. Real GDP is highest in regions with a relative high price development regardless of the liquidity cluster. Regions with a relative weaker price development combined with a strong liquidity development, i.e. cluster (2,1), are characterized as the smallest in terms of population, working population as well as real GDP. The population and working population development clearly show the demand effect on price and liquidity. The real GDP development shows a very similar picture. The strongest population increase is to observe in regions that exhibit the strongest price and liquidity increases. Hence, more productive regions attract more people what in turn leads to more demand for space and consequently results in higher prices and higher liquidity. This is the case if the supply side cannot keep up. In contrast, the “cold” markets experienced the weakest population growth. It is also to observe that population as well as working population growth is more pronounced in cluster (1,2) than in cluster (2,1). This might indicate that population growth has a stronger impact on the price development than on the liquidity development. Furthermore, the “hot” regions reveal the lowest unemployment rate as well as the highest disposable income. Especially the household disposable income exhibits a strong variation between the clusters. While cluster (1,1) displays a median disposable household income of €49.34 thousand, cluster (2,2) achieves only €42.41 thousand. However, it seems as if the poor performing regions are trying to catch up. With a decline in the unemployment rate of 36% and an increase in disposable household income of 5.95%, cluster (2,2) regions are the best performers with regard to these two variables.

**Figure 2.7** Allocation of NUTS 3 regions to price and liquidity clusters, investment market



**Notes:** The figure displays the geographic distribution of the individual clusters. The NUTS 3 regions are allocated to a specific cluster by applying the partitioning around medoids (PAM) algorithm on the price and liquidity index values. The data consists of 1,476,592 observations on the residential investment market. The sample period is 2013 Q1 to 2018 Q4.

Regarding the spatial distribution of the clusters, as shown in Figure 2.7, a strong concentration of regions allocated to cluster (1,1) and cluster (1,2) in the southern part of Germany emerges.

While for Baden-Württemberg, Bavaria and south Hesse 90%, 70% and 85% of all regions are allocated to price cluster 1, the clustering approach displays a different distribution for the northern part of the country.

On the investment market, it seems like metropolitan areas with a prosperous economic outlook backed by strong fundamental data have large spillover effects on the surrounding regions, impacting their price and liquidity development. This effect can be observed for Berlin, Frankfurt, Hamburg, Hannover and Rostock. The increased demand for ownership in regions next to metropolitan areas might be caused by the increasing unaffordability of ownership within those metropolises and their outskirts.

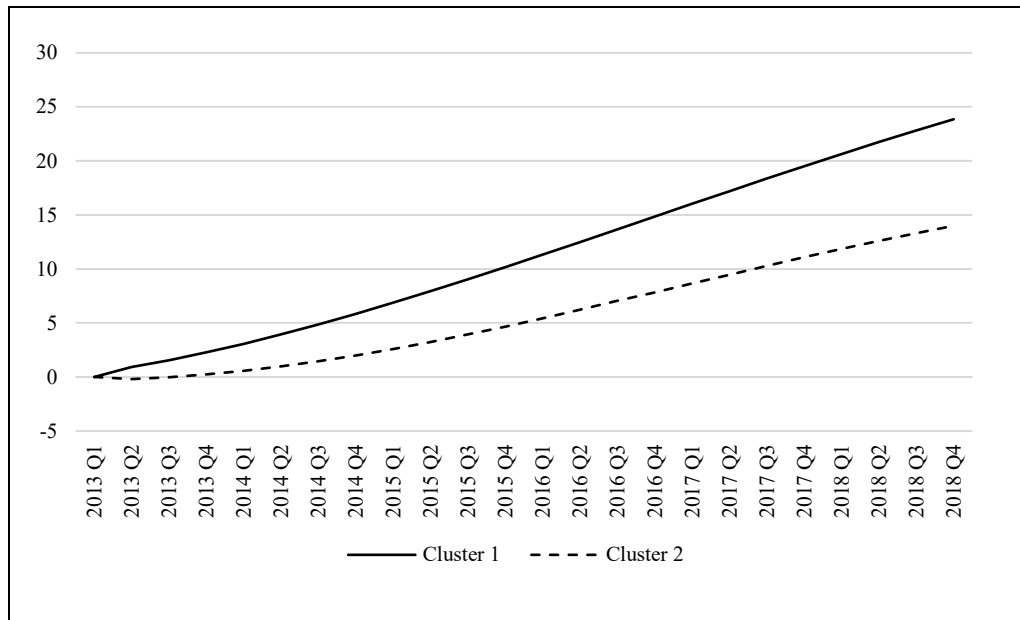
Not surprisingly the share of eastern German regions is higher in price and liquidity cluster 2. The strongest concentration of eastern German regions allocated to price and liquidity cluster 2 can be found along the Czech and the Polish border. Of all eastern German regions represented in the sample, 61% are allocated to price cluster 2, while 67% are allocated to liquidity cluster 2. A similar pattern emerges for economically underdeveloped regions as defined by the Federal Ministry for Economic Affairs and Energy (2016), based on socioeconomic data like income and unemployment and the availability of infrastructure. Of those regions 68% are allocated to price cluster 2 and 67% to liquidity cluster 2. Since a proportion but not all of those regions are located in eastern Germany, a further geographical segmentation seems worthwhile. Interestingly, a majority of the economically underdeveloped regions which are assigned to price and liquidity cluster 1 is located in eastern Germany. This might be due to the fact that the price increase in more rural areas like those around Berlin and Rostock came only after the cities themselves became unaffordable and a large number of people began to demand dwellings which did not participate in a price rally beforehand. The western regions allocated to cluster (2,2) correspond very well to the regions defined as economically underdeveloped. Those are in particular the Ruhr Area which still suffers from the decline of its heavy industry, and the remote regions around the Emsland, in the far northwest of the country, and the regions close to the Danish border.

### **2.1.5 Cluster results on the rental market**

Figure 2.8 shows the average cumulated change in quality- and spatial-adjusted prices on the rental market for the dwellings allocated to the respective price clusters. The price clusters on the rental market do not necessarily contain the same regions as the price clusters on the investment market. Price cluster 1 sets itself apart by a very strong and consistently positive development. Nevertheless, the final index value of 23.86 is even less than half of the final index value for the respective quality- and spatial-adjusted price cluster on the investment market, underlining the generally stronger increase in prices on the investment market than on the rental market. The rental market price index in price cluster 2 shows a very similar development. While it has been

slightly decreasing compared to 2013 Q1 for the following two quarters, an upward sloping trend is visible from then on. The increasing development is a lot weaker compared to price cluster 1 but nevertheless finishes with an increase of 14%.

**Figure 2.8** Price index, rental market



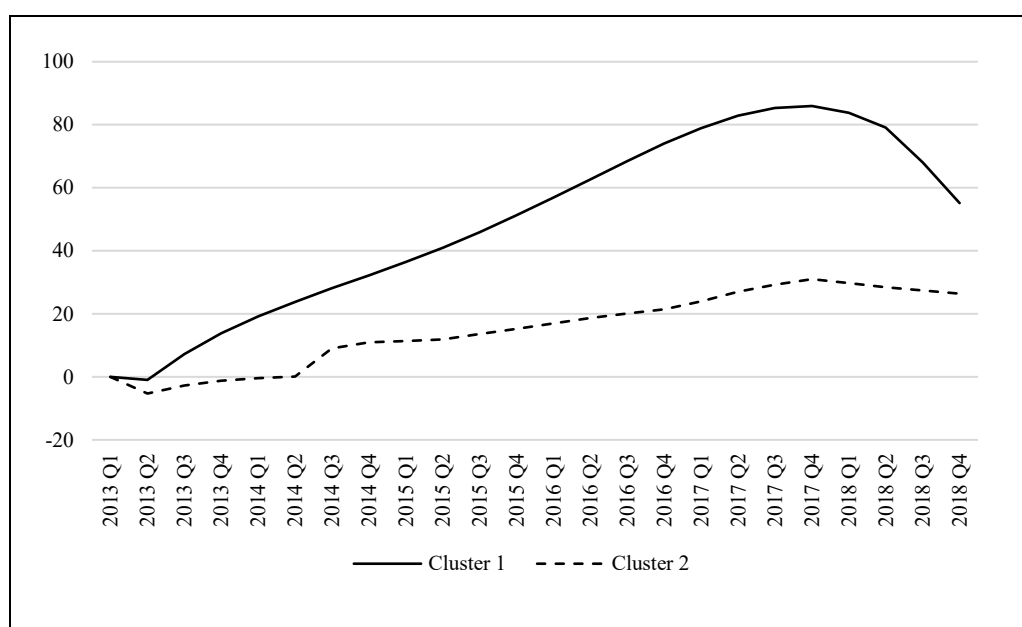
**Notes:** The figure plots the mean cumulative percentage price change for dwellings allocated to the individual price clusters. The price changes are presented as the coefficients of the time dummy variable of a quality- and spatial-adjusted GAMLSS regression. To cluster the index values, the partitioning around medoids (PAM) algorithm was used. The data consists of 3,168,458 observations on the residential rental market. The sample period is 2013 Q1 to 2018 Q4.

The liquidity development on the rental market depicted in Figure 2.9, however, clearly distinguishes itself from the investment market. After a one-period drop in the first quarter, liquidity in liquidity cluster 1 was steadily rising until it climaxed in 2017 Q4 with 85.95% above the base quarter level and afterwards fell to 55.1% at the end of the observation period. Liquidity cluster 2 decreases relative to the base quarter until 2014 Q1 and from then a slight upward trend is visible until 2017 Q4. A small decline in liquidity in liquidity cluster 2 is apparent in 2018. At the end of the observation period liquidity cluster 2 finishes with an increase of 26.39% above the baseline level. As with prices the total liquidity growth is much weaker on the rental market compared to the investment market.

The NUTS 3 regions assigned to price cluster 1 experienced a relative strong increase in prices compared to price cluster 2. At the median they are characterized by relatively low numbers with regards to population, with 143.15 thousand inhabitants in price cluster 1 compared to 169.28 thousand in price cluster 2, as well as working population. In terms of GDP, however, regions in price cluster 1 exceed the regions in price cluster 2. Thus, strong rental growth can be observed in the relatively small but economically productive regions. The difference in GDP to price cluster 2, however, is way less pronounced compared to the investment market. Regions allocated to

price cluster 1 are further characterized by lower unemployment rates and a higher disposable income of €49.14 thousand compared to €43.69 thousand in price cluster 2. A possible driver of the rental growth from 2013 Q1 to 2018 Q4 might be the rising demand for living space, stemming from the increasing population as well as working population. While population in price cluster 1 increased by 3.78%, price cluster 2 only had a growth of 1.3% and working population even declined. To summarize so far, price cluster 1 seems to obtain regions with a promising future. As on the investment market price cluster 2 experienced a slightly stronger decline in unemployment rates and a stronger increase in disposable household income. But why did liquidity evolve in such different paths on the investment and rental market in the years 2014 to 2017? Besides the increased level in building completions, the explanation might again stem from a combination of economic sentiment and interest rate expectations, which of course have a stronger impact on buying decisions. With economic sentiment declining from the start of 2014 onwards, interest rates began to fall, but so did financing expectations as defined by ZEW and JLL (2018). Although effective rates on new mortgage loans fell below 2% for the first time in early 2015<sup>10</sup>, a continuing global economic mitigation quickly deteriorated the reimbursed financing expectations. The improvement in liquidity on the investment market from 2016 Q2 onwards might have been initiated by the ECB decision to stimulate the Eurozone economy with further expansionary measures which bolstered the economy and at the same time ensured a continuing low interest rate environment.

**Figure 2.9** Liquidity index, rental market



**Notes:** The figure plots the mean cumulative percentage change in liquidity for dwellings allocated to the individual liquidity clusters. The changes are presented as the coefficients of the time dummy variable of a quality- and spatial-adjusted Cox proportional hazards model. To cluster the index values, the partitioning around medoids (PAM) algorithm was used. The data consists of 3,168,458 observations on the residential rental market. The sample period is 2013 Q1 to 2018 Q4.

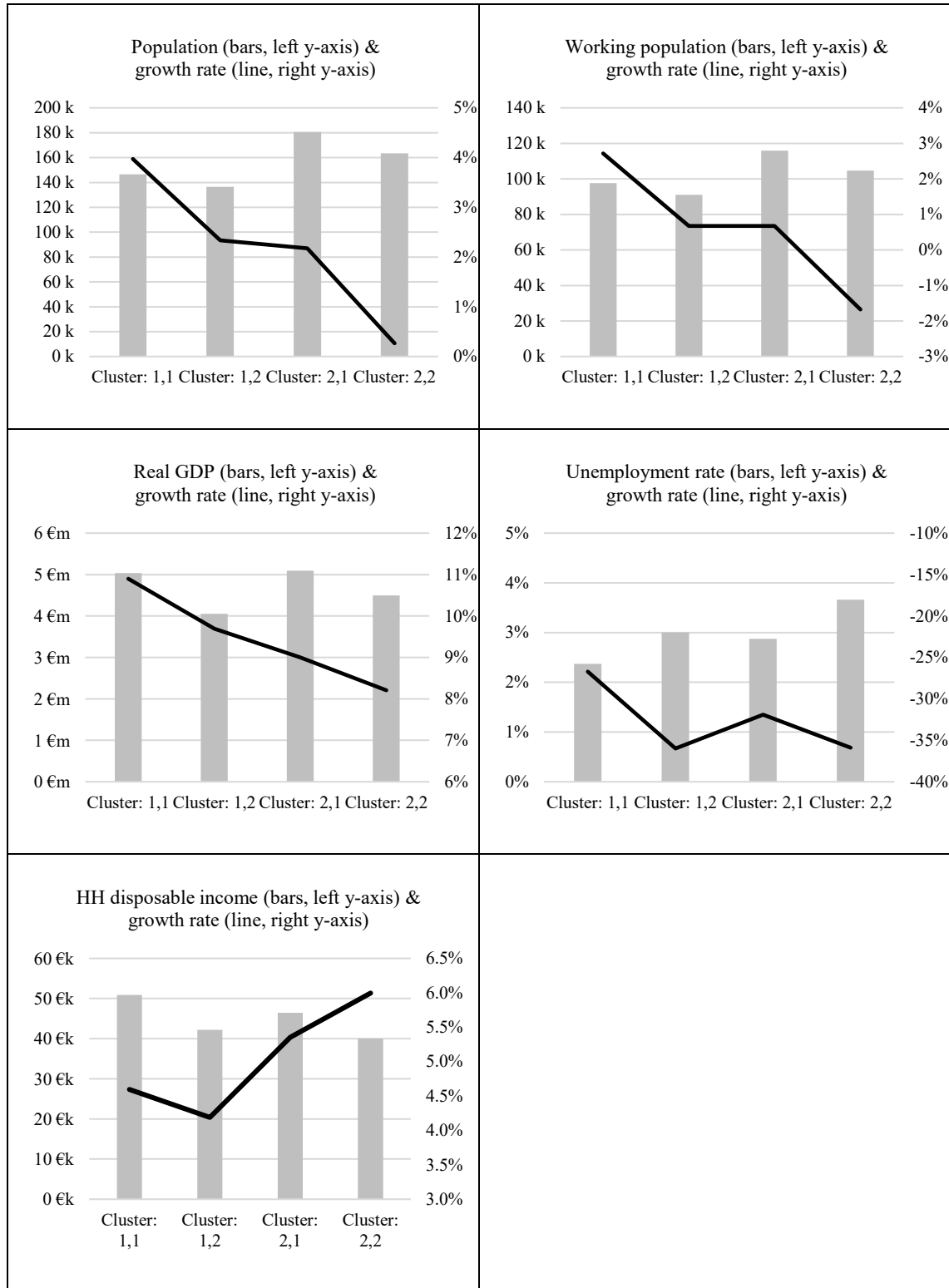
<sup>10</sup> According to Deutsche Bundesbank

Liquidity development on the rental market, regarding the economic and socioeconomic factors exhibits a very similar picture. As with prices on the rental market regions assigned to liquidity cluster 1, i.e. regions with a relatively strong liquidity development, at the median exhibit a relatively low unemployment rate of 2.6% compared to 3.58% in liquidity cluster 2 as well as a relatively high disposable household income of €48.66 thousand compared to €40.54 thousand in liquidity cluster 2. The decline in the unemployment rate was more distinct in liquidity cluster 2 and household disposable income growth was slightly stronger in liquidity cluster 2. The difference between liquidity cluster 1 and liquidity cluster 2 is very pronounced in terms of real GDP, with €5.04 million compared to €4.37 million. Thus, liquidity was especially increasing in extremely productive regions. These regions also experienced a relatively strong increase in real GDP of 10.12% in liquidity cluster 1 and 8.4% in liquidity cluster 2. With regard to population, liquidity cluster 1 and liquidity cluster 2 are very similar. Working population is slightly higher in liquidity cluster 1. The growth rates of these two variables, however, are stronger in liquidity cluster 1. While population in liquidity cluster 1 regions has been increasing by 3.22%, liquidity cluster 2 regions only experienced a small plus of 0.65%. Hence, again the increasing demand for space seems to be a reliable liquidity driver.

Out of the 158 regions assigned to cluster 1 by price, 123 regions are also assigned to liquidity cluster 1 and hence can be characterized as “hot” rental markets. These markets show the strongest development of population with an increase of 3.97%, working population with a plus of 2.72% and real GDP with a growth of 10.9% over the observation period. These factors generate a higher demand for space, leading to tight housing markets and, thus, result in a strong upward price and liquidity development on the rental market. Furthermore, these regions are characterized by the lowest unemployment rate of 2.37%, by far the highest disposable household income of €50.89 thousand and with €5.04 million real GDP is also relatively high. NUTS 3 regions assigned to the cluster combination (2,1), exhibit the highest real GDP of €5.1 million, with 180.56 thousand inhabitants are the most populated and have the highest working population of 115.95 thousand people. Therefore, demand in these regions could be relatively high, resulting in a strong liquidity development. The respective growth variables that indicate the present and future development, however, are less outstanding, leading to the relatively weak price development. With an increase of 2.33% and 9.69% respectively, cluster (1,2) exhibits the second largest increase in population and real GDP. Furthermore, with 35.98% cluster (1,2) experienced the biggest drop in the unemployment rate. These cities seem to have a promising future and, thus, the increase in demand for space leads to a pronounced increase in rents. Liquidity, however, stays weak as these markets do not seem to be too overrun. The weakest NUTS 3 regions by means of economic and socioeconomic data are the 129 regions assigned to cluster combination (2,2). They show a very low population development of 0.27% and even a loss in working population of 1.68%. Moreover, with 8.21%, 3.66% and €40.02 thousand, cluster (2,2) displays the weakest real GDP growth, the

highest unemployment rate as well as the lowest disposable household income. Due to the underlying development, strong price and liquidity increases are not expected in these regions for the near future. Figure 2.10 summarizes the economic and socioeconomic data for the four possible cluster combinations on the rental market.

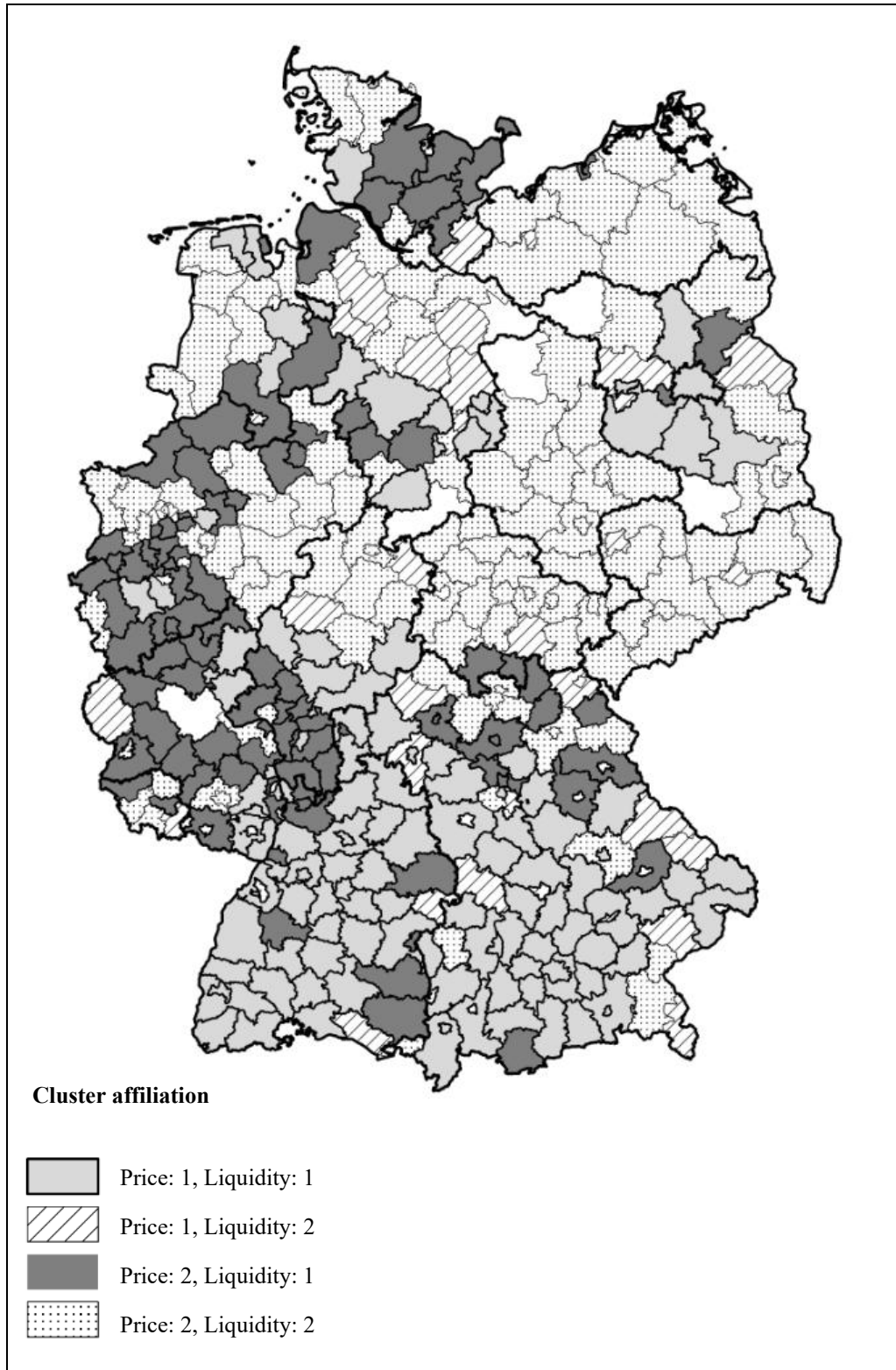
**Figure 2.10** Socioeconomic data by cluster affiliation, rental market



**Notes:** The figure displays the socioeconomic data by cluster on the rental market. The NUTS 3 regions are allocated to a specific cluster by applying the partitioning around medoids (PAM) algorithm on the price and liquidity index values, respectively. The grey bars exhibit the actual values of the respective variables, labelled on the left y-axis.

The black lines display the respective growth rates and are labelled on the right y-axis. The data consists of 3,168,458 observations on the residential rental market. The sample period is 2013 Q1 to 2018 Q4.

**Figure 2.11** Allocation of NUTS3 regions to price and liquidity clusters, rental market



**Notes:** The figure displays the geographic distribution of the individual clusters. The NUTS 3 regions are allocated to a specific cluster by applying the partitioning around medoids (PAM) algorithm on the price and liquidity index values. The data consists of 3,168,458 observations on the residential rental market. The sample period is 2013 Q1 to 2018 Q4.

As depicted in Figure 2.11, the geographic analysis of the rental market displays a more differentiated pattern. Again, most of the strongest markets, in terms of price and liquidity development, are located in the southern part of Germany. In the northern part, however, almost exclusively the economically very strong performing regions are allocated to price and liquidity cluster 1. Among those are for example Berlin, Bremen, Cologne, Hanover and Wolfsburg, where the Volkswagen AG has its headquarters and a large production facility.

Within the more differentiated pattern, the allocation of eastern German regions is also much more descriptive. While the share of eastern German regions allocated to price and liquidity cluster 2 for the investment market was between 60 and 70%, on the rental market 83.3% of all eastern German regions in the sample are allocated to price cluster 2 and 86.1% to liquidity cluster 2. The preference for western German regions on the rental market is emphasized by comparing the ratio of the share of eastern German regions allocated to cluster (1,1) to the share of those allocated to cluster (2,2). For the investment market, the ratio of the share of eastern German regions allocated to the strongest cluster in relation to the weakest cluster is 1:2, while on the rental market the ratio is 1:10. So the share of eastern German regions allocated to the “coldest” markets is much higher on the rental market. The same holds true for the economically underdeveloped regions of which 75.8% are allocated to price cluster 2 and 64.7% (slightly less than on the investment market) to liquidity cluster 2. While on the investment market the distribution of economically weak regions to the clusters is indifferent of the location of those regions, on the rental market 84% of the economically underdeveloped regions assigned to price cluster 1 are not in eastern Germany. For the liquidity cluster 1 it is 83%.

These findings imply that on the rental market the demand for dwellings in economically weak regions is much less pronounced than on the investment market. As economically non-performing regions are not in the centre of attention of developers and landlords, we neglect increased supply in this context. This very selective demand for economically solid regions illustrates one of the largest benefits of the rental market, which is the higher flexibility tenants have in comparison to owners. For tenants it is much easier to select their place of residence or to form a decision to relocate based on fundamental data. In a polycentric country like Germany which has at the same time comparatively high ancillary acquisition costs for real estate of up to 10%, a large rental market facilitates a more flexible domestic migration.

### **2.1.6 Summary and further implications**

With its very granular analysis of 380 regions and the subsequent classification of the regions by price and the corresponding market liquidity on the residential investment and rental market, the study was able to reveal new insights to the German residential market and to draw valuable implications.

1. Over the same observation period and in the same regions, the ratio of dwellings offered for rent to those offered for investment was 2:1, demonstrating the size and importance of the rental market in Germany.
2. The optimal way to analyse commonalities and differences between the investment market and the rental market of 380 German NUTS 3 regions is by allocating the regions to two clusters based on their price and liquidity development, respectively.
3. On the investment market, the allocation to the price cluster is strongly depending on the actual values of fundamental data like population, working population and real GDP, as well as their growth rates. The allocation to liquidity clusters is much more depending on growth rates in those variables, i.e. liquidity comes with economic improvement.
4. The geographic analysis revealed that most of the regions allocated to price and liquidity cluster 1 are located in the economically strong southern half of the country. In addition, it seems that on the investment market, strong economic hubs have large spillover effects on surrounding regions.
5. On the rental market, regions allocated to price cluster 1 exhibit a relatively small size in terms of population and working population but show strong economic fundamentals like GDP, unemployment rate and disposable household income. For the liquidity clustering the size of the regions plays an important role as well.
6. The geographic analysis for the rental market reinforces these findings, as a very strong dependence on economically strong regions is revealed.
7. Regarding the four analysed categories, price and liquidity on the investment and rental market respectively, it is obvious that regions allocated to cluster 1 are characterized by relatively strong and favourable economic and socioeconomic conditions and developments in recent periods.

As stated above, the strong demand for economically prosperous regions on the rental market is not very surprising and exhibits one of the most important benefits of renting a dwelling, which is flexibility. If the quality of the flat or the chosen region does not live up to the expectations or worsened since moving in, both can be substituted within a timespan of three month. In addition to that, a large and functioning rental market enables people willing to move without having the financial resources to buy a home. A circumstance which is non-negligible for people moving to more prosperous economic regions.

The strong demand for economically prosperous regions on the investment market might only be obvious at first sight. The reader with a background in economics might suppose that buying a dwelling only in prosperous regions is the most natural thing to do, however, buying a home in Germany is more regarded as consumption than as an investment. Professional landlords on the other hand should of course take the fundamentals into account, when doing the market analysis before buying new dwellings. This is where another peculiarity of the German real estate market

comes in. Only a small proportion of dwellings offered for rent is actually offered by profit oriented private institutions, whereas the majority is offered by non-professional individual landlords. According to the GdW (2016) approximately 37% of all dwellings are offered by non-professional landlords whereas only about 8% are offered by commercial institutions. Since the dwellings which experienced the strongest price increase seem to be chosen in a very institutional way by analysing fundamental data, a possible implication is, that the broad mass of private non-professional landlords how they are called, actually became very professional and sophisticated real estate investors. So in contrast to the stock market, for which German individuals are known to not be very attracted to, the residential real estate market is where they quite professionally invest their money.

While the above was the rather positive message of this article, the ongoing price increase on both the investment and the rental market further decreases the affordability of living space. This is a very topical issue in Germany, as especially in cities and metropolitan regions, large parts of the population are not able to create ownership or have to accept a monthly rent at or even above 40% of the monthly household income. Whereas this study is not able to derive specific strategies for providing affordable living space or easing tight residential markets, it might assist policy makers in detecting overheating markets and facilitate a categorization of regions according to their market tightness. Herewith, policy makers should have a better understanding of which regions they should focus on and where market interventions are most urgent. The separate clustering enables policy makers to derive regionally varying strategies based on the pricing of markets, the tightness of markets measured by liquidity, and a combination of both.

## **Conclusion**

It is the aim of this paper to build quality- and spatial-adjusted price indices for the major German residential investment and rental markets on a regional basis and to complement each with a liquidity index, in order to obtain a very detailed assessment of the German residential real estate market. In the context of this study, liquidity is examined with a time-based measure and does not include transaction cost, volume-based, price impact or return-based measures.

While the mere analysis of descriptive statistics on the investment and rental market already indicated a trend of rising prices and declining time on market, the analysis of quality- and spatial-adjusted price indices eliminates the possibility that these trends are due to changes in the quality of the housing stock. Therefore a cluster analysis is conducted, separately for the price and liquidity development on the investment and rental market. For each of these four categories the average silhouette method suggests the optimal number of clusters to be two. The regions are then assigned to one of two clusters based on the partitioning around medoids (PAM) clustering

algorithm, in order to identify common market movements and to facilitate the interpretation of the results for the 380 regions over the observation period from 2013 Q1 to 2018 Q4.

The highest price increase on the investment market is identified for dwellings within the economically strongest and largest regions, which also exhibit higher growth rates of the variables depicting the demand for space. The liquidity development is strongest in regions that are relatively small in terms of population and working population, however, have experienced the largest growth rates in the socio-economic variables. While most of the regions allocated to the clusters with the sharpest increase in price and liquidity are found in the southern half of the country, economically strong metropolitan regions in the northern part of the country seem to have a significant impact on the price and liquidity development of the surrounding regions. Those spillover effects are revealed for cities like Berlin, Frankfurt, Hamburg, Hanover and Rostock.

On the rental market, the dwellings which experienced the sharpest price increase are found in the economically strongest, however, relatively small regions. The allocation to the highest liquidity cluster can be found for economically very strong regions supported by relatively high population growth. The geographic analysis underlines the importance of strong economic data for significant demand on the rental market, since most eastern German regions as well as the economically underdeveloped regions are assigned to the weaker price and liquidity clusters.

To summarize, it becomes obvious that the allocation to cluster 1 for price as well as liquidity on the investment and rental market is supported by higher growth rates in the variables population, working population, and real GDP, implying higher demand for space. Furthermore, in each of the four categories cluster 1 reveals a lower unemployment rate as well as a higher disposable income. The decline in the unemployment rate as well as the increase in the disposable household income, however, is more pronounced in cluster 2. This might stem from the relatively high or respectively low values these variables in cluster 2 come from.

One of the most interesting implications which can be drawn based on the findings of this paper is, that apparently a large part of the German population has developed into professional real estate investors. In Germany, the largest share of landlords is the one of the so-called non-professional landlords. As the regions assigned to the cluster with the most significant price increase seem to be chosen based on a very sophisticated market analysis, by identifying the regions with the strongest fundamental data, it seems like the dominating market players have significantly increased their knowledge and approach for investing in residential real estate.

A peculiarity of the analysis described in the paper is, that it is easily applicable to other residential markets. The findings of this paper are of course limited by the rather short sample period of only six years. Price and liquidity development before 2013 Q1 could not be incorporated because of the absence of data. Furthermore, due to a lack of transaction prices on the entire German residential real estate market asking prices operate as a “take it or leave it option” to the buyer or tenant. While this assumption seems to be plausible on the rental market, the availability of

transaction prices on the investment market could further improve the knowledge of and the insights about the German residential real estate market. This topic will certainly be of major interest for future research as soon as data availability in Germany becomes less restricted. Moreover, an analysis of the intertemporal relationship between the investment and rental market, incorporating time lags, might be of further interest in order to gain deeper insights to the interdependencies of the investment and rental market. This classification could then be used to extend the price versus rent literature. The center of a city might be specified not only geographically but rather as an economic and sociodemographic city center. Furthermore, the proximity to local public transport and other amenities, tourism hotspots, commuting flows as well as the distance to adjacent urban areas might improve the insights of this analysis. The type of neighborhood, schools and noise, e.g. from a highway, should be considered as well. Including these additional spatial variables, would lead to a more precise and detailed spatial mapping of a city. These variables would most likely lead to a change in the impact of the current variable “distance to the NUTS 3 center”.

## References

- an de Meulen P, Mitze T (2014) Exploring the spatial variation in quality-adjusted rental prices and identifying hot spots in Berlin's residential property market. *Regional Studies, Regional Science* 1:310-328
- Allen MT, Rutherford RC, Thomson TA (2009) Residential asking rents and time on the market. *Journal of Real Estate Finance and Economics* 38:351-365
- Ametefe F, Devaney S, Marcato G (2011) Estimating liquidity in real estate markets. IMF Research Program, 2015
- Amihud Y, Mendelson H (1986) Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17:223-249
- Amihud Y (2002) Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5:31-56
- Bauer TK, Feuerschütte S, Kiefer M, an de Meulen P, Micheli M, Schmidt T, Wilke LH (2013) Ein hedonischer Immobilienpreisindex auf Basis von Internetdaten: 2007–2011. *ASTA Wirtschafts- und Sozialstatistisches Archiv* 7:5-30
- Berkovec JA, Goodman JL Jr (1996) Turnover as a measure of demand for existing homes. *Real Estate Economics* 24:421-440
- Brounen D, Eichholtz P, Ling D (2009) The liquidity of property shares: an international comparison. *Journal of Real Estate Economics* 37:413-445
- Cajias M (2018) Is there room for another hedonic model? The advantages of the GAMLSS approach in real estate research. *Journal of European Real Estate Research* 11:204-245
- Cauley SD, Pavlov AD (2002) Rational delays: the case of real estate. *Journal of Real Estate Finance and Economics* 24:143-165
- Clayton J, MacKinnon G (2000) Measuring and explaining changes in REIT liquidity: moving beyond the bid-ask spread. *Real Estate Economics* 28:89-115
- Collett D, Lizieri C, Ward C (2003) Timing and the holding periods of institutional real estate. *Real Estate Economics* 31:205-222
- Cox DR (1972) Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34:187-220
- Das S, Hanouna P (2010) Run lengths and liquidity. *Annals of Operations Research* 176:127-152
- Demsetz H (1968) The cost of transacting. *Quarterly Journal of Economics* 82:33-53
- Deutsche Bundesbank (2020) Effektivzinssätze Banken DE / Neugeschäft / besicherte Wohnungsbaukredite an private Haushalte (Code BBK01.SUD531). Accessed 28 March 2020

Diewert EW (2003) Hedonic regressions: A review of some unresolved issues. 7th Meeting of the Ottawa Group, Paris, May

Federal Ministry for Economic Affairs and Energy: Koordinierungsausschuss der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ (GRW) (2016) Regionalpolitischer Bericht der Bund-Länder-Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur“ 2016. [https://www.bmwi.de/Redaktion/DE/Publikationen/Wirtschaft/regional\\_politischer-bericht\\_bund-laender-gemeinschaftsaufgabe-verbesserung-regionale-wirtschaftsstruktur16.pdf?\\_\\_blob=publicationFile&v=4](https://www.bmwi.de/Redaktion/DE/Publikationen/Wirtschaft/regional_politischer-bericht_bund-laender-gemeinschaftsaufgabe-verbesserung-regionale-wirtschaftsstruktur16.pdf?__blob=publicationFile&v=4). Accessed 16 August 2019

Federal Statistical Office (2019a) Baufertigstellungen von Wohn- und Nichtwohngebäuden unter anderem nach Bauherren – Lange Reihen. <https://www-genesis.destatis.de/genesis/online> (Code 31121-0001). Accessed 17 August 2019

Federal Statistical Office (2019b) Preisindizes für Wohnimmobilien. <https://www-genesis.destatis.de/genesis/online> (Code 61262-0001). Accessed 17 August 2019

Federal Statistical Office (2019c) Verbraucherpreisindex für Deutschland. <https://www-genesis.destatis.de/genesis/online> (Code 61111-0003). Accessed 17 August 2019

Fisher J, Gatzlaff D, Geltner D, Haurin D (2003) Controlling for the impact of variable liquidity in commercial real estate price indices. *Real Estate Economics* 31:269-303

Fisher J, Ling D, Naranjo A (2009) Institutional capital flows and return dynamics in private commercial real estate markets. *Real Estate Economics* 37:85-116

Follain JR, Velz OT (1995) Incorporating the number of existing home sales into a structural model of the market for owner-occupied housing. *Journal of Housing Economics* 4:93-117

GdW Bundesverband deutscher Wohnungs- und Immobilienunternehmen e.V. (2016) Anbieterstruktur auf dem deutschen Wohnungsmarkt. <https://web.gdw.de/uploads/pdf/infografiken/2016/Anbieterstruktur.pdf>. Accessed 17 August 2019

Geltner M, Miller N, Clayton J, Eichholtz P (2014) *Commercial real estate analysis and investments*. 3<sup>rd</sup> edition, Ohio OnCourse Learning, Mason

Genesove D, Mayer CJ (1997) Equity and time to sale in the real estate market. *American Economic Review* 87:255-269

Glascock J, Lu-Andrews R (2013) An examination of macroeconomic effects on the liquidity of REITs. *Journal of Real Estate Finance and Economics* 49:23-46

Goldberger AS (1968) The interpretation and estimation of Cobb-Douglas functions. *Econometrica: Journal of the Econometric Society* 36:464-472

de Haan J (2010) Hedonic price indexes: a comparison of imputation, time dummy and ‘re-pricing’ methods. *Jahrbücher für Nationalökonomie und Statistik* 230:772-791

- Halvorsen R, Pollakowski HO (1981) Choice of functional form for hedonic price equations. *Journal of Urban Economics* 10:37-49
- Heinrich M, Just T (2016) *Wie heterogen sind deutsche Städte? Zur Segmentierung deutscher Wohnimmobilienmärkte. Schriftenreihe der Universität Regensburg: Beiträge zur Immobilienwirtschaft*
- Held T, Waltersbacher M (2015) *Wohnungsmarktprognose 2030. BBSR-Analysen KOMPAKT, Nr. 7. Bundesamt für Bauwesen und Raumordnung, Bonn*
- Just T, Heinrich M, Wiersma S (2019) *Segmentierung deutscher Wohnimmobilienmärkte, gif e.V. working paper*
- Kain JF, Quigley JM (1970) Measuring the value of housing quality. *Journal of the American Statistical Association* 65:532-548
- Karolyi GA, Lee KH, Van Dijk MA (2012) Understanding commonality in liquidity around the world. *Journal of Financial Economics* 105:82–112
- Kaufman L, Rousseeuw PJ (1987) Clustering by means of medoids. In: Dodge Y (ed) *Statistical data analysis based on the L1-norm and related methods*, Elsevier/North-Holland, Amsterdam, pp 405-416
- Kaufman L, Rousseeuw PJ (1990) *Finding groups in data: an introduction to cluster analysis*. Wiley, New York
- Kluger BD, Miller NG (1990) Measuring residential real estate liquidity. *Real Estate Economics* 18:145-159
- Krainer J (1999) Real estate liquidity. *Economic Review* 3:14-26
- Lancaster KJ (1966) A new approach to consumer theory. *Journal of Political Economy* 74:132-157
- an de Meulen P, Mitze T (2014) Exploring the spatial variation in quality-adjusted rental prices and identifying hot spots in Berlin's residential property market. *Regional Studies, Regional Science* 1:310-328
- Michels W, Hiller N, Oberst C (2011) *Wohnungsmarktregionen in Deutschland: Abgrenzung und Struktur funktionaler Wohnungsmärkte. Institut für Siedlungs- und Wohnungswesen*
- Pastor L, Stambaugh RF (2003) Liquidity risk and expected stock returns. *Journal of Political Economy* 111:642–685
- Rigby RA, Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54:507-554

- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20:53-65
- Sarr A, Lybek T (2002) Measuring liquidity in financial markets. IMF Working Paper No. 02/232
- Shimizu C, Nishimura KG, Watanabe T (2012) House prices from magazines, realtors, and the land registry. In: Bank of International Settlements (ed), *Property Market and Financial Stability*, BIS Papers No. 64, pp 29-38
- Stein JC (1995) Prices and trading volume in the housing market: A model with down-payment effects. *Quarterly Journal of Economics* 110:379-406
- Syed I, Hill RJ, Melser D (2008) Flexible spatial and temporal hedonic price indexes for housing in the presence of missing data. UNSW Australian School of Business Research Paper
- Triplett J (2004) Handbook on hedonic indexes and quality adjustments in price indexes: application to information technology products. STI OECD Working Papers No. 09
- Voigtländer M (2009) Why is the German homeownership rate so low? *Housing Studies* 24:335-372
- Zentrum für Europäische Wirtschaftsforschung (ZEW) GmbH and Jones Lang LaSalle SE (2018) DIFI-Report 3. Quartal 2018. <http://ftp.zew.de/pub/zew-docs/DIFI/DIFI092018.pdf>. Accessed 28 March 2019

### **3 Exploring the determinants of real estate liquidity from an alternative perspective: censored quantile regression in real estate research**

#### **Abstract**

In this paper, the liquidity (inverse of time on market) of rental dwellings and its determinants for different liquidity quantiles are examined for the seven largest German cities. The determinants are estimated using censored quantile regressions in order to investigate the impact on very liquid to very illiquid dwellings. As market heterogeneity is not only observed between cities but also within a city, each of the seven cities is considered individually. Micro data for almost 500,000 observations from 2013 to 2017 is used to examine the time on market. Substantial differences in the magnitude and direction of the regression coefficients for the different liquidity quantiles are found. Furthermore, both the magnitude and direction of the impact of an explanatory variable on the liquidity, differ between the cities. To the best of the authors' knowledge this is the first paper, to apply censored quantile regressions to liquidity analysis of the real estate rental market. The model reveals that the proportionality assumption underlying the Cox proportional hazards model cannot be confirmed for all variables across all cities, but for most of them.

**Acknowledgement:** The authors especially thank PATRIZIA Immobilien AG for contributing the dataset for this study.

All statements of opinion are those of the authors and do not necessarily reflect the opinions of PATRIZIA Immobilien AG or its associated companies.

## Introduction

The concept of asset liquidity in the residential rental market is somewhat blurred. In the investment market, asset liquidity traditionally measures the time it takes the owner to turn an asset into cash. In the rental market on the other hand, asset liquidity measures the time it takes the landlord to find a new tenant, i.e. from introducing the dwelling to the market until the signing of the rental contract. In this paper, liquidity is defined as the inverse of the time on market (TOM) of rental dwellings, i.e. the higher the time on market, the lower the liquidity.<sup>11</sup> Whether the letting process is quick or slow, depends mainly, on the initial asking rent, the structural quality and location of the asset, demand for space, and the overall market conditions. A detailed understanding of these major drivers of marketing time of rental dwellings is the objective of this paper. To extract more insightful information from the available data, the study introduces a modelling technique which is new to the field and uses the results to validate the outcome of the most established approach to measure liquidity, which is the Cox (1972) proportional hazards model (PHM).

This paper focuses on the residential rental markets of the seven largest German cities (descending order by population): Berlin, Hamburg, Munich, Cologne, Frankfurt, Stuttgart, and Dusseldorf. While for the methodological aim of the paper, these markets are subsidiary, it is useful to understand the German rental regulatory framework and gain some insights into the housing markets in order to follow the contextual aim of this specific study. By law, tenants in Germany usually have a three months cancelation period, for which reason a dwelling is typically brought onto the market before the tenant leaves.<sup>12</sup> With a national ownership rate of roughly 43% as of 2013, the first year of the sample period, clearly more than half of German households rent their homes and hence the rental market deserves closer attention. Voigtländer (2009), Bentzien et al. (2012), Lerbs and Oberst (2014), and Reisenbichler (2016) explain in detail the reasons for this distinctive market feature. The considered cities exhibit ownership rates that are far lower than the German average, ranging from 33% in Stuttgart to 16% in Berlin. Therefore, especially in these cities, the rental market plays a significant role. While six of the cities can be generalized as long-established markets, Berlin shows stronger growth rates in population, purchasing power, employment, rents and purchase prices for real estate, however, starting from lower absolute values in those figures. For decades, the western German cities have developed into trans-regional hubs for education and employment. Berlin has “only” been able to benefit from the nationwide improvement after the reunification of Germany in 1990, however, since then its positive development was substantial.

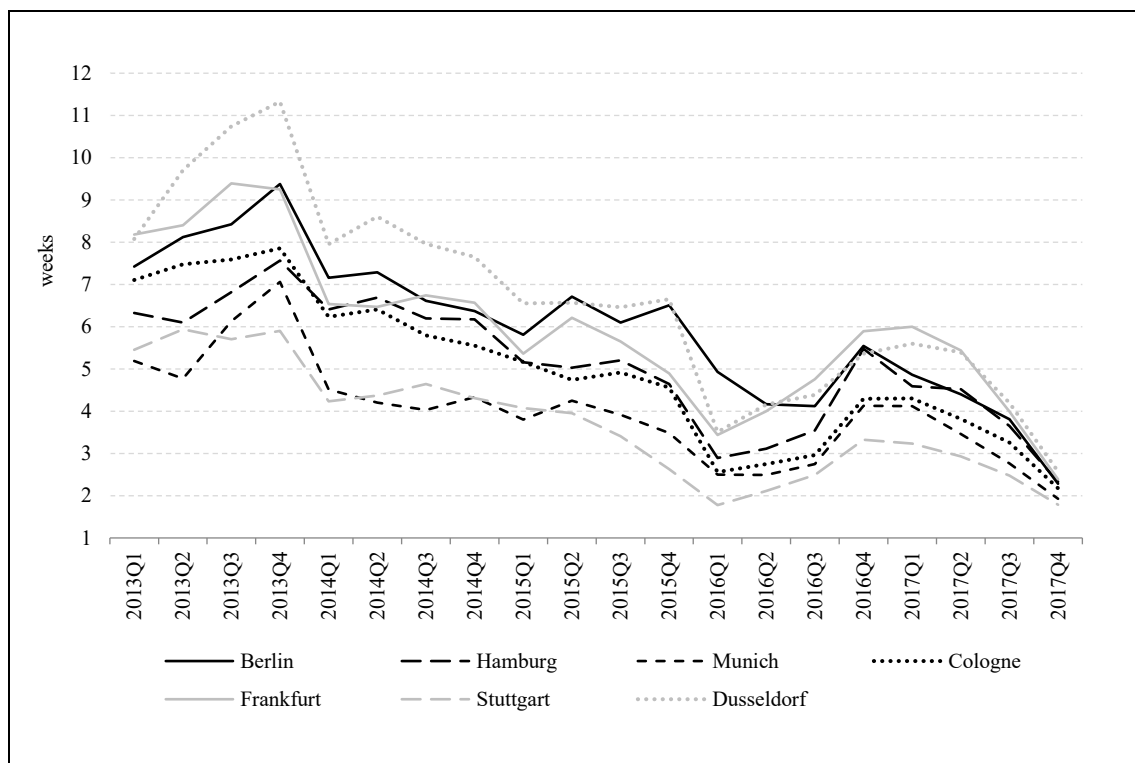
---

<sup>11</sup> Based on the definition of Wood and Wood (1985).

<sup>12</sup> Landlords can only terminate the contract to use the dwelling for themselves or for other household members.

Figure 3.1 illustrates the average time on market for the seven largest German cities extracted from the underlying dataset. Despite some up and down movements, the graph clearly shows a continuous decline in time on market within the last years. This development points to an increasingly strong demand on the rental market, as prospective tenants have to shorten their decision-making process because of high competition for an insufficient supply. This excess demand was recognized by Held and Waltersbacher (2015) who claimed that this comprises 272,000 newly constructed dwellings per year on a national level for the years 2015-2020, while the actual completion did not satisfy this demand.<sup>13</sup> In addition, the use of online platforms might lead to better informed market participants, resulting in a narrower price assessment and hence less market uncertainty and increased market efficiency.

**Figure 3.1** Average time on market in weeks in the seven largest German cities



**Notes:** This figure displays the average time on market in weeks in the seven largest German cities. The data consists of 482,196 observations of the residential rental market. The sample period is 2013 Q1 to 2017 Q4.

Since the aim of this paper is to understand the drivers of demand for rental space by means of a well-established method and the introduction of a method new to the field of real estate, the first step is to explore time on market via the Cox PHM. Next, an advanced econometric approach is introduced. To the best of the authors' knowledge this study is the first, to investigate the determinants of time on market by applying a censored quantile regression (CQR) in real estate

<sup>13</sup> According to the Federal Statistical Office (2018) dwelling completion was: 216,727 (2015), 235,658 (2016) and 245,304 (2017).

The CQR aims to explain the variation of time on market as a function of dwelling characteristics and other spatial and socioeconomic characteristics. The decisive feature of the analysis is that CQRs are used to model any quantile of the distribution of the dependent variable. Chaudhuri et al. (1997) stress this feature as a great advantage compared to mean regressions, as distributions might not only be different in terms of their means but might differ especially in their upper and lower tails. CQRs can quantify the impact of a covariate on the dependent variable for any quantile, compared to only the center of the population. The CQR, as an expansion of the survival regression analysis, is expected to yield more accurate estimations and to provide a more complete statistical analysis for understanding the factors driving liquidity and the underlying demand. The dataset consists of 482,196 observations on the rental markets of the seven largest German cities between 2013 Q1 and 2017 Q4. The study reveals that the magnitude and direction of impact of an explanatory variable on time on market differs between the cities. Hence, this implies the importance of analyzing each city individually and not the seven cities or even the German market as whole. Furthermore, the magnitude and direction of effect of an explanatory variable on the liquidity of a dwelling exhibits differences between time on market quantiles within a city. This implies that, the proportional hazards assumption, underlying the Cox PHM is violated for individual explanatory variables and cities and thus emphasizes the use of the CQR approach for the time on market analysis. The study concludes that the heterogeneity across the liquidity quantiles, as well as the heterogeneity between the cities, are accountable for the distinguishable impacts of changes in the covariates on time on market. These findings should of course be of interest to current and future landlords, as they reveal both the characteristics of dwellings along the liquidity distribution (e.g. the existence of built-in kitchen), as well as the impact of a change in characteristics on the liquidity of dwellings (e.g. impact of installing a built-in kitchen). Therefore, landlords should be able to infer whether a dwelling displays the characteristics of a highly liquid thus highly demanded dwelling, or what actions they could take in order to increase the expected liquidity, e.g. install a built-in kitchen or change the floor plan to increase the number of rooms. Furthermore, the findings suggest that nationwide or even statewide policy measures might not be expedient to address the specific situation on the residential rental real estate market of a particular region, city or neighborhood.

The remainder of this paper proceeds as follows: The next Section provides a literature review. Section 3.3 describes the underlying econometric model, followed by a detailed description of the dataset and the descriptive statistics in Section 3.4. Estimation results are presented and discussed in Section 3.5. Section 3.6 concludes.

## Literature review

Belkin et al. (1976) conducted one of the first empirical studies analyzing real estate liquidity for different market segments. They divide the market according to geographic areas, price segments and buyers' search space. By doing so, they analyze the relationship between time on market and the spread between listing price and selling price, using ordinary least squares (OLS) estimation techniques. They find essential differences between market segments. Especially in high-price segments, deviations from the initial list price had a more pronounced effect on time on market. The determinants of time on market considering different price segments have comprehensively been analyzed in the literature. Kang and Gardner (1989) found that the impacts on time on market do vary in magnitude between low-, medium-, and high-price segments. While Kang and Gardner (1989) did not identify the simultaneity problem between time on market and the selling price, Yavas and Yang (1995) applied a two-stage least squares (2SLS) estimation to deal with the fact that time on market and price mutually influence each other. They exhibit a significant positive impact of price on time on market in the medium-price segment, whereas this effect is insignificant for houses in the low- and high-price segments. Allen et al. (2009) also use a multi-step procedure to analyze the relationship between asking rent and time on market in the rental market for single-family residential rental listings. Based on asking rents, the sample is divided into three price segments. They find that underpricing of asking rents and time on market move in the same direction in all price segments, although the effect is stronger in the medium- and high-price segment. In further studies real estate markets have been segmented according to the number of rooms, the number of units in a structure, the geographical region, the property type or by market cycle, respectively. The conclusion emerging from these findings is that market segmentation seems to make a valuable contribution to our understanding of liquidity patterns.

More closely related to the present study is the article of Turnbull and Dombrow (2006), given that they divide their sample into low-, medium- and high-liquidity segments. They explore the impact of listing density on time on market for a pooled sample, for different market cycles and for different market cycles combined with different liquidity segments. Applying a three-stage least squares (3SLS) estimation, they find that the significance, as well as the magnitude and directions of the impact of spatial competition variables on time on market, vary between the different liquidity segments.

Based on these findings, the following hypotheses are deduced and investigated in this paper:

1. Difference between quantiles: The direction and magnitude of the effect of covariates on real estate liquidity are not equal for and vary across low, medium and high-liquidity segments. If this is the case, the assumption of proportional hazards underlying the Cox PHM would be violated, justifying the need for an approach able to deal with heterogeneous effects.

2. Difference between cities: It is hypothesized that the direction and magnitude of the impact of the covariates on time on market vary across the seven cities. If this is true, an approach jointly capturing the residential rental markets of the seven cities or even regarding a national market might lead to the loss of important information.

Nowadays, the most popular model for the estimation of duration data is the Cox PHM. Also commonly used is the accelerated failure time (AFT) model. However, in econometric terms, this paper differs substantially from the preceding studies, as CQRs are applied to real estate liquidity analysis. Quantile regression (QR) has been formally introduced by Koenker and Bassett Jr. (1978). Compared to the accelerated failure time model or the Cox PHM, QR is a more flexible estimation method, as it allows for consistent estimation of the regression model without restrictions on the variation of estimated coefficients over the quantiles. The decisive feature of the analysis, however, is that QRs are used to model any quantile of the distribution of the dependent variable. Chaudhuri et al. (1997) stress this feature as a great advantage compared to mean regressions, as distributions might not only be different in terms of their means but might differ especially in their upper and lower tails. Thus, QRs can quantify the impact of a covariate on the dependent variable for any quantile, compared to only the center of the population. In contrast to linear regression, QR coefficients are computed via minimizing the sum of weighted absolute deviations.

Since its introduction, the QR approach has received increasing attention, theoretically as well as empirically, and has been applied to many different research areas.<sup>14</sup> In the real estate literature, more precisely in the area of hedonic pricing, QRs have been applied by Zietz et al. (2008) and Liao and Wang (2012), among others. Thomschke (2015) for example used the method for the German market. However, when it comes to real estate liquidity, this is the first paper, to the best of the authors' knowledge, to use QRs with censoring for duration analysis on the real estate market. For the closely related analysis of (un)employment durations, Horowitz and Neumann (1987) initially, as well as Lüdemann et al. (2006), Schmillen and Möller (2012) among others, have lately applied CQRs. Conceptually the present analysis is highly related to Lüdemann et al. (2006). In particular, the CQR method used in this paper goes back to Koenker and Biliias (2002).

---

<sup>14</sup> For survival analysis, see e.g. Crowley and Hu (1977), Koenker and Geling (2001); for medical research, see e.g. Beyerlein et al. (2008), Wehby et al. (2009); for financial economics, see e.g. Bassett Jr. and Hsiu-Lang (2002); for environmental research, see e.g. Hendricks and Koenker (1992); for labour economics, see e.g. Buchinsky (1994, 1995).

## Econometric approach

### 3.1.1 Cox proportional hazards model

Without doubt, the leading model for the analysis of survival data is the Cox PHM. This model is used for exploring the determinants of the duration of an event or elapse of time. For example, it determines the variables that accelerate or restrict the elapse of time that a response variable needs to change its state. In this case, the response variable is defined as a non-negative continuous variable, measuring the time elapse that a dwelling requires to change its status from being offered on the market to being out of the market in weeks, i.e. time on market. For understanding and estimating survival data, two main functions are necessary: the survival function  $S(t)$  and the hazard rate function  $\lambda(t)$ . The survival function specifies the probability that an event has not occurred until a certain time  $t$  and is formally defined as

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^{\infty} f(x)dx, \quad (1)$$

with  $f(x)$  being the probability density function (p.d.f.) of the time until the event. The hazard function  $\lambda(t)$ , by contrast, describes the probability at  $t$  that an event occurs at time  $T$ , given that the event has not occurred before and is given by

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | t \leq T)}{\Delta t}. \quad (2)$$

The survival function expresses the probability of a dwelling staying on the market, while the hazard function measures the risk of the same dwelling leaving the market. The Cox regression for a specific observation  $i$  is given as

$$\lambda_i(t|x_i) = \lambda_0(t) \exp(-x_i' \beta), \quad (3)$$

where  $x$  is a vector of covariates (without the constant),  $\beta$  is a vector of parameters and  $\lambda_0(t)$  is the non-negative baseline hazard. The Cox PHM requires no specification of the functional form of the baseline hazard  $\lambda_0(t)$ . However, it assumes proportional hazards, meaning that the hazard function is a constant function of time. The elapse of time that a dwelling is offered on the market corresponds to an event that might be right-censored. Right-censoring arises when the landlord does not change the status of the dwelling in the Multiple Listing Services (MLS) database, or the dwelling is still being offered on the market. The Cox regression framework allows the censored events of the sample to contribute to the model until the end of the observation period. Therefore, a semiparametric PHM is estimated for each of the  $k=7$  cities according to

$$h(t) = \exp(X\beta) + u. \quad (4)$$

The hazard function  $h$  of the time on market  $t$  depends on the covariate matrix  $X$  plus an iid error  $u$ .

### 3.1.2 Cox proportional hazards model versus censored quantile regression

While the Cox PHM is the most common tool for explaining time on market in social sciences, natural sciences and real estate studies, new techniques have been developed to account for conditional survival functions across different levels of response. The traditional Cox regression estimates the conditional survival function for the entire sample, based on the assumption of homoscedasticity within the sample. The covariates are expected to exert the same impact on the response, regardless of the distribution of the response. In contrast, the quantile approach estimates different coefficients for different quantiles of the population. A traditional example when explaining quantile regression is the duration of unemployment. When using a Cox PHM, the effect of the covariate *experience* will not distinguish between long-term and short-term unemployed persons. In contrast, the CQR takes the different quantiles of the response, i.e. long-term and short-term unemployment into consideration and estimates several equations with different elasticities.

In this context, quantile regressions have arisen as a method for estimating conditional regressions within the sample as a function of the quantile distribution of the response. For the analysis of duration data CQRs are estimated to overcome the issue of right-censoring. This paper is the first to introduce the CQR to the context of real estate liquidity, to the best of the authors' knowledge. The Cox PHM has an advantage over the CQR regarding the issues of competing risks, time-varying covariates and unobserved heterogeneity. However, the CQR does not come with the cost of the proportional hazards assumption that needs to be empirically proven. Hence, misspecification of the model is less likely. The model requires only very weak assumptions on the error term, for example homoskedasticity of the error terms is not required, and thus is more robust against non-normal errors, outliers and misspecification of the error term. Moreover, Powell (1986) demonstrates that under appropriate conditions for a certain value of  $\tau$ , the censored regression quantile estimator  $\beta^\tau$  is consistent and asymptotic normality is proven, if the appropriate assumptions hold for each  $\tau \in \{\tau_1, \dots, \tau_J\}$ , irrespective of the distribution of the error term. Furthermore, the interpretation of regression results is straight forward. In summary, the CQR yields a robust and more flexible alternative for the estimation of parametric and semiparametric duration models and provides a very comprehensive statistical analysis.

### 3.1.3 Quantile regression model

The origin of the QR model extends back to Koenker and Bassett Jr. (1978). It is a location model estimating the relationship between the covariate matrix  $X$  and the dependent variable  $y$ , conditional on the quantile  $\tau$  of  $y$ . The quantile  $\tau \in (0, 1)$  is defined as the value of  $y$  that separates the observations into the fraction  $\tau$  below and the fraction  $1 - \tau$  above.

Applying the log-transformation of the time on market  $T_i$  of a particular dwelling  $i$ , according for example, to Chaudhuri et al. (1997), yields the accelerated failure time model as basis for the relationship between time on market and the covariates dependent on the conditional quantile. The underlying model can be described as

$$\ln T_i = x_i' \beta^\tau + u_i^\tau. \quad (5)$$

The conditional quantile functions of the logarithm of the time on market can be written as

$$\text{Quant}_\tau(\ln T_i | x_i) = x_i' \beta^\tau, \quad (6)$$

where  $\text{Quant}_\tau(\ln T_i | x_i)$  represents the  $\tau$ th conditional quantile of  $\ln T_i$  given a  $k \times 1$  vector of covariates  $x_i$ . The conditional quantile of the iid error term  $u$ ,  $\text{Quant}_\tau(u^\tau | x)$ , is 0.

### 3.1.4 Censored quantile regression model

An important feature of survival analysis is that some observations do not change their event status throughout the observation period, as some dwellings remain available in the MLS database until the end of the observation period. If this is the case, the response variable, time on market  $T_i$ , is right-censored. To deal with censoring within the QR framework, three main approaches have been introduced by Powell (1984, 1986), Portnoy (2003) and Peng and Huang (2008). For the present dataset, Powell's (1984, 1986) approach, which addresses fixed censoring, is applied. For CQRs with fixed censoring, it is necessary to know the observation-specific censoring value  $C_i$  for all observations. If an observation  $i$  is censored, it is not possible to observe the actual survival time  $T_i$ , but only to observe the observation specific censoring value  $C_i$  instead. Thus, in a right-censored dataset  $T_i$  is given by  $T_i = \min \{T_i^*, C_i\}$ .  $C_i = \ln T_i$ , if an observation is censored and  $C_i = +\infty$ , if an observation is not censored. The CQR estimator  $\widehat{\beta}^\tau$  is the value of  $\beta^\tau$  solving the minimization problem of the distance function

$$Q_N(\beta; \tau) \equiv \frac{1}{N} \sum_{i=1}^N \rho_\tau(\ln T_i - \min(x_i' \beta^\tau, C_i)). \quad (7)$$

Thus,  $x_i' \beta^\tau$  is censored from above at the upper threshold  $C_i$ . The ‘‘check-function’’  $\rho_\tau(u)$  is defined as:

$$\rho_\tau(u) = \begin{cases} \tau * |u| & u \geq 0 \\ (1 - \tau) * |u| & u < 0. \end{cases} \quad (8)$$

$\tau * |u|$  denotes the penalty for underprediction and  $(1 - \tau) * |u|$  for overprediction. The estimator  $\widehat{\beta}$  that minimizes the distance function  $Q_N(\beta; 0.5)$ , i.e. at the median  $\tau = 0.5$ , describes a special case yielding the censored least absolute deviations (LAD) estimator  $\widehat{\beta}^{0.5}$ . The coefficients can be interpreted as the change in the dependent variable that, ceteris paribus, arises from a marginal change in the respective regressor, while keeping the dependent variable in the same quantile, according to Machado and Mata (2000). An increase in an explanatory variable by a marginal

unit, *ceteris paribus*, prolongs or shortens the time on market by  $[|1 - \exp(\widehat{\beta^\tau})| * 100]\%$  keeping time on market in the same quantile. A prolongation of the time to event occurs if the respective coefficient  $\beta^\tau$  is positive, i.e. the hazard ratio  $\exp(\widehat{\beta^\tau})$  is greater than 1, and a reduction of the time to event if  $\beta^\tau$  is negative, i.e. the hazard ratio is smaller than 1. For example, if  $\beta_{refurbished}^{0.2}$  was 0.077, the interpretation would be as follows: Dwellings of the 0.2 quantile which are newly renovated, *ceteris paribus*, stay 8.1% longer on the market, than dwellings which are not newly renovated.

## Data and descriptive statistics

The estimation sample is composed of three merged datasets, containing information from 482,196 observations of single- and multi-family rental dwellings in the seven largest German cities from the first quarter of 2013 to the fourth quarter of 2017. Observations regarding student housing, affordable housing, temporary housing and retirement housing have not been incorporated into the dataset. The variables, their units and sources can be found in the Online Appendix in Table A1. Information on the rental dwellings was provided from empirica systems ([www.empirica-systeme.de](http://www.empirica-systeme.de)) which collects georeferenced real estate data from more than 100 German Multiple Listing Systems (MLS) such as ImmoScout, Immonet or Immowelt but also regionally focused marketplaces and newspapers for the whole German market. As the market leader of real estate data for Germany, empirica has an own proprietary algorithm that identifies duplicates and harmonizes the sample of big data. The variable of major interest, time on market, is defined as the number of weeks a dwelling was listed in the MLS calculated by the start and end date according to Benefield and Hardin (2015) for example. The asking rent is defined in absolute terms measured in euros per month. The common issue of reverse causality between time on market and rent, discussed by Yavas and Yang (1995) among others, is not addressed in this paper due to a lack of contract rents on the German residential market. In this paper the asking rent is set by the landlord at the beginning of the data generating process (DGP). The time on market is established as the time the dwelling is offered in the MLS given the initially set asking rent. The asking rent operates as a “take it or leave it option” to the tenant and thus rent negotiations are not considered. On the residential rental market this assumption is plausible as negotiations about the monthly rental payments, especially in the overheated markets of the seven major cities considered in this paper, are rather an exception. Hence, this approach provides a descriptive analysis, however, is not able to draw causal conclusions. Since the data is georeferenced, two spatial gravity indicators, measuring the Euclidian distance of each dwelling to the geographical centroid of the ZIP and NUTS 3 polygon in kilometers, are included. NUTS 3 regions correspond to “the nomenclature of territorial units for statistics”, which is a hierarchical system for dividing the economic territory in Europe. The NUTS 3 regions cover small regions

similar to counties or administrative districts. In the sample, each city represents one NUTS 3 region and therefore, the distance to the NUTS 3 centroid describes the distance to the geometric city center. The socioeconomic variables of purchasing power per household and the number of households at the ZIP code level, are obtained from the “Gesellschaft für Konsumforschung” (GfK). The population density per km<sup>2</sup> in a ZIP code area is calculated in ArcGIS. The final source is Thomson Reuters Eikon, providing the 10-year interest rate for housing loans as a macro variable. The analysis focusses on seven time on market quantiles from the 0.2 to the 0.8 quantile. The top and bottom end of the time on market distribution display extremely high and extremely low values respectively and hence are not the subject of this analysis.

Table 3.1 shows the descriptive statistics for the variables of interest for each of the seven cities. Comparing the seven cities shows that the mean time on market is fairly low in Stuttgart and Munich and relatively high in Dusseldorf, Berlin and Frankfurt. With 3.73 weeks, the mean time on market is lowest in Stuttgart and highest in Dusseldorf with 6.6 weeks. Table 3.2 shows the time on market for each of the selected quantiles and for all seven cities. It confirms that the time on market is lowest in Stuttgart and Munich and highest in Dusseldorf, Frankfurt and Berlin along the entire distribution. Furthermore, Table 3.2 exhibits the degree of censoring present in the data for each city. The degree of censoring can be interpreted as the percentage of dwellings that have not been taken out of the MLS until the cut of date. For example, in Berlin 6.7% of all dwellings that appeared in the MLS from 2013 Q1 to 2017 Q4 still remain in the MLS after the last day of the observation period.

The covariate of asking rent is lowest in Berlin with a mean of 692.37 € and highest in Munich with a mean of 1,209.59 €. Interestingly, these cities respectively exhibit the lowest and highest mean purchasing power of 35,272.76 € in Berlin and 55,942.79 € in Munich. Furthermore, they are characterized having on average the oldest and youngest buildings respectively. This indicates high development activity in the more recent past in Munich. These variables reveal interesting differences between the cities and give rise to an individual investigation of each city.

The aim of the study is to investigate the impact of changes in the explanatory variables on the time on market of rental dwellings, divided by their liquidity level. Hence, it is of special interest whether there are patterns in the dwelling characteristics, which might explain the affiliation to a respective liquidity quantile. Figure 3.2 shows that across all seven cities, the dwellings in the most liquid quantile, 0.2 (Q20), are on average the cheapest, the smallest and the oldest. Furthermore, they are located in ZIP code areas with the lowest purchasing power, but a relatively large number of households. In contrast to the most liquid quantile, dwellings assigned to the 0.8-quantile (Q80) display on average 33.71% higher rents, are 25.3% larger, 12.13% younger, located in ZIP codes with 2.67% higher purchasing power, have 3.46% less households in a ZIP code area and are 3.05% more densely populated. These insights based on pure analysis of descriptive statistics give rise to the assumption that the level of time on market strongly depends

on the characteristics of a dwelling. Hence, this supports the CQR approach to get a deeper understanding of the determinants of time on market.

**Table 3.1** Descriptive statistics of selected variables

<b>Panel A: Berlin, ownership rate: 15.61%</b>					
<b>Variable</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Dev.</b>	<b>Q1</b>	<b>Q99</b>
Asking rent	692.37	549.94	470.12	233.42	2,500.35
Time on market	5.95	2.60	8.71	0.10	55.50
Living area	74.27	67.00	34.08	29.50	200.01
Age	65.57	59.00	39.20	0	117.00
Households in ZIP	12,008	11,997	3,645.74	3,890	20,434
Purchasing power of HHs in ZIP	35,272.76	34,352.80	4,954.13	27,548.31	52,669.38
Population density	3,899	3,542	2,398.80	61	7,908
N=180,858					
<b>Panel B: Hamburg, ownership rate: 21.14%</b>					
<b>Variable</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Dev.</b>	<b>Q1</b>	<b>Q99</b>
Asking rent	795.44	658.55	483.79	254.11	2,624.01
Time on market	4.98	2.10	7.51	0.10	40.20
Living area	71.49	65.65	30.30	26.00	180.00
Age	52.12	52.00	34.56	0	117.00
Households in ZIP	11,051	10,677	3,458.73	2,143	17,979
Purchasing power of HHs in ZIP	43,559.36	42,779.87	7,670.23	32,723.47	63,894.32
Population density	3,800	3,757	2,342.76	45	7,560
N=101,008					
<b>Panel C: Munich, ownership rate: 25.23%</b>					
<b>Variable</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Dev.</b>	<b>Q1</b>	<b>Q99</b>
Asking rent	1,209.59	1,034.08	721.71	379.96	3,960.00
Time on market	3.84	1.70	6.30	0.10	33.60
Living area	77.66	72.00	37.20	23.00	209.33
Age	42.22	41.00	33.69	0	117.00
Households in ZIP	11,459	12,074	3,241.20	3,573	16,896
Purchasing power of HHs in ZIP	55,942.79	54,728.80	6,132.63	45,586.18	69,752.31
Population density	4,173	4,463	2,184.26	253	7,933
N=47,394					
<b>Panel D: Cologne, ownership rate: 27.42%</b>					
<b>Variable</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Dev.</b>	<b>Q1</b>	<b>Q99</b>
Asking rent	720.55	639.6	369.88	250.00	2,039.92
Time on market	5.04	2.40	7.28	0.10	39.47
Living area	72.03	68.00	30.13	23.00	168.97

Age	45.24	46.00	29.60	1.00	117.00
Households in ZIP	13,452	13,521	3,594.13	6,176	20,561
Purchasing power of HHs in ZIP	45,466.38	44,370.20	5,748.89	34,685.48	58,827.02
Population density	3,675	3,390	2,243.07	395	7,598

N=47,527

**Panel E: Frankfurt, ownership rate: 20.67%**

Variable	Mean	Median	Std. Dev.	Q1	Q99
Asking rent	1,012.81	850.20	634.55	299.98	3,499.85
Time on market	5.89	2.70	8.29	0.10	45.90
Living area	79.03	72.00	37.23	23.00	208.00
Age	49.63	47.00	39.57	0	117.00
Households in ZIP	11,147	11,669	4,351.39	1,546	20,945
Purchasing power of HHs in ZIP	47,528.31	46,692.74	6,663.02	37,419.27	76,088.03
Population density	3,930	4,194	2,098.80	146	7,785

N=41,446

**Panel F: Stuttgart, ownership rate: 32.92%**

Variable	Mean	Median	Std. Dev.	Q1	Q99
Asking rent	910.71	775.20	501.26	270.98	2,749.82
Time on market	3.73	1.60	5.93	0.10	30.40
Living area	79.19	74.00	34.29	23.00	193.01
Age	50.25	48.00	34.85	0	117.00
Households in ZIP	10,385	10,927	3,098.97	1,104	15,899
Purchasing power of HHs in ZIP	47,058.05	46,374.77	4,440.57	40,041.62	61,972.64
Population density	3,507	3,353	1,766.72	254	7,404

N=17,967

**Panel G: Dusseldorf, ownership rate: 24.08%**

Variable	Mean	Median	Std. Dev.	Q1	Q99
Asking rent	762.86	630.00	486.69	240.00	2,579.00
Time on market	6.60	3.20	8.91	0.10	50.01
Living area	75.65	70.00	33.85	25.00	190.00
Age	52.18	54.00	29.78	1.00	117.00
Households in ZIP	9,726	9,703	3,021.54	2,721	15,045
Purchasing power of HHs in ZIP	47,869.27	46,140.99	5,851.67	40,382.03	65,472.45
Population density	3,999	3,913	2,370.53	23	7,906

N = 45,996

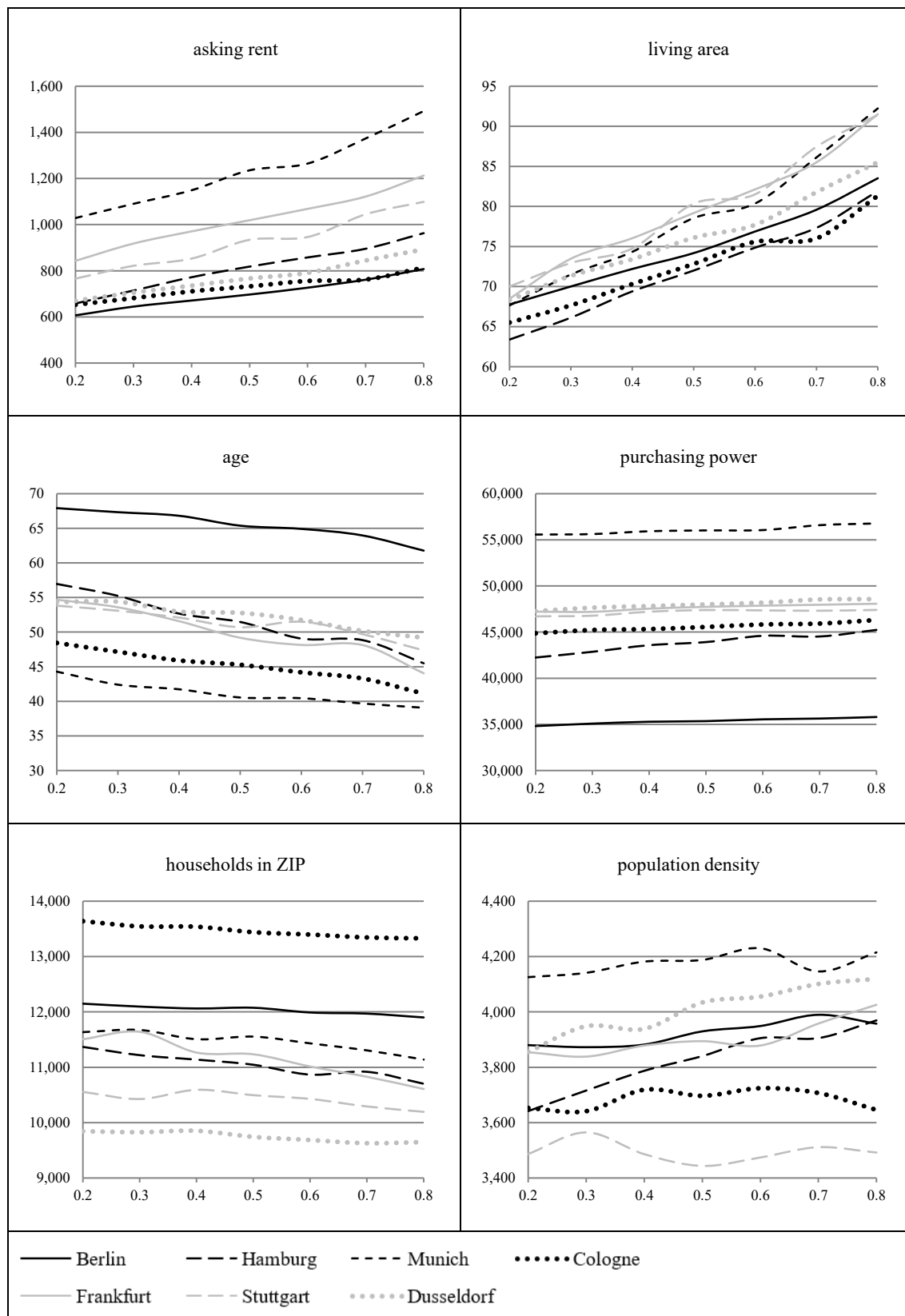
**Notes:** This table reports selected descriptive statistics, the ownership rate and the number of observations for each of the seven cities. The ownership rates are as of 2013. The data consists of 482,196 observations on the residential rental market. The sample period is 2013 Q1 to 2017 Q4.

**Table 3.2** Degree of censoring and average time on market in weeks per quantile

	Degree of Censoring	Q20	Q30	Q40	Median	Q60	Q70	Q80
Berlin	6.7%	0.6	1.1	1.7	2.6	3.8	5.6	8.8
Hamburg	6.0%	0.5	0.9	1.4	2.1	3.1	4.7	7.4
Munich	5.2%	0.4	0.7	1.1	1.7	2.4	3.5	5.3
Cologne	4.9%	0.6	1.1	1.6	2.4	3.4	4.8	7.4
Frankfurt	6.7%	0.7	1.2	1.9	2.7	4.0	5.7	8.7
Stuttgart	3.7%	0.3	0.7	1.1	1.6	2.3	3.5	5.4
Dusseldorf	7.0%	0.9	1.5	2.2	3.2	4.6	6.4	10.0

**Notes:** This table reports the degree of censoring and the average time on market in weeks for the seven cities. The data consists of 482,196 observations on the residential rental market. The sample period is 2013 Q1 to 2017 Q4.

**Figure 3.2** Descriptive development of selected covariates for time on market quantiles



**Notes:** This figure displays descriptive statistics for selected variables across seven time on market quantiles for the seven largest German cities. The time on market quantiles are labelled on the x-axis and the respective diagram title is labelled on the y-axis. Altogether, the data consists of 482,196 observations on the residential rental market. The sample period is 2013 Q1 to 2017 Q4.

## Results

### 3.1.5 Results of the cox survival regression

In a first step, covariates raising or lowering the time on market of rental dwellings of the seven largest German cities have been considered. The results of the Cox survival regressions are presented in Table 3.3. Since the results are displayed as hazard rates, a rate larger than one decreases the time on market and thus increases liquidity, while a rate smaller than 1 decreases liquidity. The results show that an increase in the asking rent *ceteris paribus*, leads to a longer time on market across all seven cities. The same direction of the effect was found by Cajias and Freudenreich (2018). The seller faces the trade-off between the time the dwelling is on the market and the price he might receive, as mentioned by Anglin et al. (2003) for the real estate transaction market. According to that, the result is not surprising, that a higher asking rent leads to lower liquidity of dwellings. Since the densely populated cities regularly display an excess demand for housing, it is of particular interest to investigate the rental effect for individual liquidity quantiles, as the magnitude of these effects is supposed to differ widely along the distribution. Miller (1978) postulates that housing attributes, defining its attractiveness, are important determinants of time on market. An increase in living area significantly increases liquidity in two of seven cities, namely Berlin and Hamburg. As Hamburg is the city with the smallest average living area in this sample, people might desire larger apartments and hence take them off the market faster. Cologne is the only city where living area has a prolonging impact on time on market, however, is not significant. A segmentation into liquidity quantiles might be useful in order to consider the impact and its significance of, for instance, the living area for each quantile, as it might be the heterogeneity within the cities that leads to insignificant effects. The number of rooms in a dwelling, in contrast, has a statistically significant positive effect on liquidity for all seven cities. Hence, it seems that in most of these seven highly demanded regions, a larger dwelling in general is not favorable. A dwelling with a larger number of rooms, however, can provide more living space for more people and thus is highly demanded and consequently off the market more quickly. Surprisingly, the marketing time is *ceteris paribus* shorter, the older the dwelling, and is longer for newly built and refurbished ones. In Berlin the marketing time is longer the older the building, however, the effect is not statistically significant. This might be due to the fact, that in this sample Berlin exhibits by far the highest average building age. An increase in the distance to the NUTS 3 centroid, which is used as a proxy for the city center, is associated with a longer time on market. This seems obvious, as regions farer away from the city center are less demanded. The distance to the ZIP code centroid shows almost no statistically significant impact. The coefficients of each socioeconomic factor are not significant for most cities, again emphasizing the importance of considering each time on market quantile separately. A higher purchasing power results on average in a shorter time on market in Berlin, Hamburg, and Dusseldorf, all else equal, whereas for the remaining cities, the effect is not statistically significant. Among the seven cities, Berlin

and Hamburg are the cities with the lowest average purchasing power, hence this result points to the fact that in Berlin and Hamburg people would spend additional purchasing power on the search and matching process that would shorten marketing time. While an increase in the number of households significantly reduces time on market in Frankfurt and Dusseldorf, the population density has a significant prolonging effect on time on market in Hamburg, Munich, and Frankfurt. It should be noted that neither socioeconomic variable has an impact on time on market in Cologne and Stuttgart. The Pseudo- $R^2$  based on Kendall's tau, which measures the concordance between estimated survival time and the observed survival time for the non-censored response sample, ranges from 64.3% to 67.5%. Those values are common in survival studies. At this point it is necessary to note that these values cannot be interpreted directly or compared to the usual  $R^2$  calculated for OLS regressions.

**Table 3.3** Results of the Cox survival regression

<b>Hazard rates and robust standard errors</b>	Berlin	Hamburg	Munich	Cologne	Frankfurt	Stuttgart	Dusseldorf
<b>Housing characteristics – metric</b>							
Log asking rent in €	0.285 0.052***	0.332 0.059***	0.396 0.058***	0.429 0.059***	0.378 0.095***	0.379 0.095***	0.410 0.058***
Log living area	1.329 0.053***	1.185 0.060***	1.035 0.061	0.994 0.062	1.026 0.079	1.151 0.102	1.039 0.070
Number of rooms	1.172 0.012***	1.165 0.012***	1.122 0.012***	1.134 0.015***	1.158 0.014***	1.096 0.019***	1.160 0.013***
Age	0.998 0.002	1.003 0.002	1.008 0.003***	1.007 0.003**	1.007 0.002***	1.002 0.003	1.003 0.003
<b>Housing characteristics – binary</b>							
With bathtub	0.873 0.016***	0.964 0.013***	0.961 0.012***	0.970 0.018*	0.906 0.016***	0.939 0.020***	0.903 0.018***
With built-in kitchen	1.102 0.015***	0.964 0.018**	1.047 0.016***	1.120 0.015***	1.138 0.014***	1.093 0.033***	1.178 0.013***
With parking slot	0.917 0.020***	0.892 0.017***	0.936 0.016***	0.982 0.024	0.919 0.018***	0.988 0.030	0.989 0.020
With terrace	0.938 0.014***	0.989 0.017	0.933 0.016***	0.940 0.021***	0.922 0.025***	0.983 0.026	0.984 0.026
With balcony	1.002 0.014	1.047 0.017***	1.003 0.014	1.009 0.021	1.005 0.017	0.989 0.020	1.016 0.023
With elevator	0.974 0.022	0.916 0.019***	1.035 0.018**	0.852 0.023***	0.955 0.022***	0.913 0.025***	0.942 0.020***
Newly built	0.830 0.025***	0.922 0.020***	0.839 0.021***	0.940 0.028**	0.921 0.021***	0.917 0.030***	0.922 0.026***
Refurbished	0.983 0.013	0.907 0.014***	0.967 0.014**	0.987 0.014	0.938 0.017***	0.960 0.021*	0.930 0.012***
<b>Spatial variables</b>							
Longitude	0.683 0.132***	0.597 0.236***	1.346 0.184	0.226 0.296***	1.310 0.426*	0.443 0.308***	0.075 0.404***
Latitude	0.679	1.447	1.300	0.218	1.824	0.806	0.106

	0.262***	0.314	0.412	0.477***	0.499**	0.521	0.700***
Distance to ZIP centroid	0.961	0.922	0.973	0.976	0.903	0.976	0.945
	0.030*	0.027***	0.029	0.027	0.046	0.033*	0.037
Distance to NUTS 3 centroid	0.942	0.952	0.982	0.949	0.964	0.975	0.971
	0.004***	0.006***	0.005***	0.010***	0.009***	0.008***	0.008***
<b>Socioeconomic variables at ZIP level</b>							
Log purchasing power	1.781	1.281	0.854	1.166	1.135	0.811	1.526
	0.126***	0.126**	0.142	0.216	0.181	0.156	0.187**
Log number of households	1.009	1.016	0.994	0.944	1.057	1.045	1.071
	0.038	0.059	0.040	0.074	0.027**	0.033	0.037*
Log population density	1.001	0.945	0.969	0.965	0.959	0.983	0.991
	0.014	0.016***	0.014**	0.023	0.018**	0.018	0.014
<b>Financial conditions at day of release</b>							
Effective 10Y IR for housing loan	1.118	1.039	0.943	1.083	0.954	0.829	1.164
	0.101	0.075	0.129	0.100	0.121	0.174	0.112
<b>Fixed effects</b>							
Construction dummies	Included						
Quarterly dummies	Included						
Intercept	Included						
Spatial adjusted (Lin-Wei) standard-errors	Considered						
R <sup>2</sup> -concordance	65.4%	66.7%	67.5%	64.3%	65.4%	66.7%	64.4%

**Notes:** This table displays the hazard ratios and robust standard errors of a Cox PHM of the time on market of a dwelling in weeks for housing, spatial and socioeconomic variables, as well as the effective ten-year interest rate and various fixed effects. In order to obtain robust standard errors, the Lin-Wei (Lin and Wei 1989) procedure, similar to the sandwich covariance matrix of the OLS estimator, is applied and extended by a spatial component, i.e. the ZIP codes. The Pseudo-R<sup>2</sup> based on Kendall's tau measures the concordance between estimated survival time and the observed survival time for only the non-censored response sample. The data consists of 482,196 observations on the residential rental market. The sample period is 2013 Q1 to 2017 Q4. \*Significant at the 10%-level; \*\* significant at the 5%-level; \*\*\* significant at the 1%-level.

### 3.1.6 Results of the censored quantile regression

In a second step, the same regressors as for the Cox survival regressions are used to estimate CQRs, in order to gain deeper insight into the time on market quantiles. Therefore, for each city, the sample of dwellings was divided into time on market quantiles from the 0.2- to the 0.8-quantile. The results for the covariates of interest are shown in Figures 3.3 to 3.6. In Figures 3.4

to 3.6 the black lines plot the estimated coefficients, the grey lines display the standard deviations. Each plot shows the development of a coefficient  $\beta_k^\tau$  over the liquidity quantiles  $\tau$  for each of the seven cities  $k$ . A positive statistically significant coefficient increases time on market, thus decreasing liquidity, while a negative statistically significant coefficient has the opposite effect. The main effects, divided into quarterly factors, housing characteristics, spatial gravity variables, and socioeconomic characteristics, are reported below. The crucial point is that, contrary to the traditional regression models, the effect of a change in the covariate holds for the same quantile  $\tau$ , rather than across quantiles.<sup>15</sup>

### 3.5.2.1. Quarterly time effects

The considered period is characterized by low interest rates, high migration to Germany, especially to the metropolises, and additionally far too low housing supply in these cities. Consequently, vacancy has mostly been diminishing, and real estate prices as well as rents increasing. Despite rising construction activity in most cities, building completion was insufficient to meet demand, leading to excess demand. Time fixed effects as quarterly dummies have been included in the estimation equation to capture these developments over time and its impact on the time on market.<sup>16</sup> This time trend can be observed in Figure 3.3 for each city and for quantiles representing high, medium and low liquidity. The base quarter is the first quarter of 2013, so that all changes are with respect to this basis. All in all, time on market has been decreasing compared to the base quarter for all time on market quantiles and for each city. However, the magnitude and direction of change differs between the cities and different time on market quantiles. In the first quarter of the year, time on the market is usually relatively low for rental dwellings. Hence, in 2013 and partly in the beginning of 2014 time on market has been increasing relative to the base quarter in all cities and all time on market quantiles. Over the course of 2014 time on market started to decrease relative to 2013 Q1 in six of the seven cities and all time on market quantiles. The exception is Hamburg, where time on market started to decrease relative to 2013 Q1 later than in all other cities and it is striking that the date of decline is documented later for more liquid quantiles. This supports the importance of analyzing all cities separately. Furthermore, it emphasizes the hypothesis that the time on market development within a city might be different for distinct time on market quantiles and thus the quantiles should be analyzed separately. By far the most pronounced decline in time on market experienced Berlin in the 0.5 and 0.8 quantile. In the 0.2 quantile the development in Berlin is rather average compared to the other cities as a sharp decline only starts in 2017. Hence, it seems that in Berlin relative to

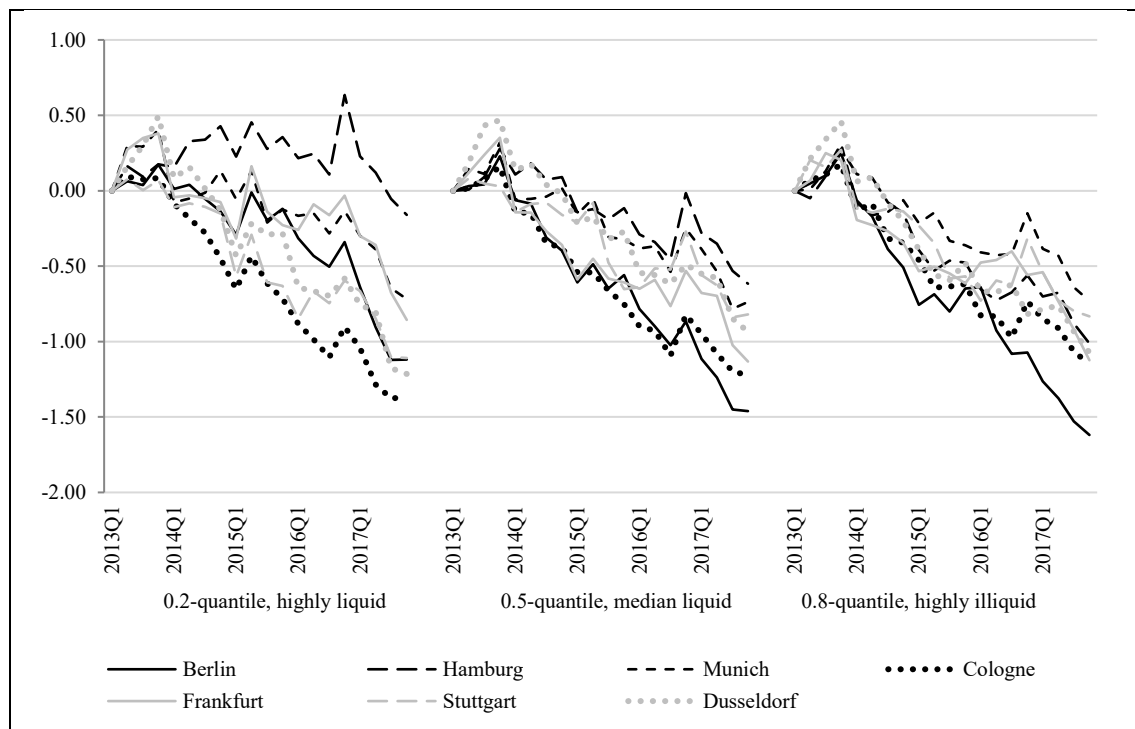
---

<sup>15</sup> Table A2 in the Online Appendix shows the estimated coefficients for all cities and quantiles.

<sup>16</sup> Of those quarterly effects, 82.0% are significant at the 10% significance level, while 74.5% are significant at the 5% significance level.

2013 Q1 the demand for rental dwellings was higher in the 0.5 and 0.8 quantile. As described in Section 3.4 Data and Descriptive Statistics, the living area of dwellings in these quantiles is relatively larger than in the 0.2 quantile. A reason for this result might be the development of household members during the recent years. According to Vonovia and CBRE (2016), in 2014, the year when time on market started to decline compared to the base quarter, Berlin was amongst the cities with the highest percentage of single households and even had the highest percentage of one- and two-person-households together. Hence, the demand for larger dwellings was probably relatively small compared to the other cities. According to data provided from Oxford Economics, over the observation period from 2013 to 2018 Berlin experienced an increase in population of 6.83%. Working population, however, has only increased by 4.94%. This is by far the most pronounced divergence of population and working population among the analyzed cities. It might be that relative to the other cities especially families with members younger than the working age, who demand relatively more living space, have moved to Berlin. Consequently, the demand and thus liquidity in Berlin have especially increased in the upper liquidity quantiles. Furthermore, during the observation period Berlin has experienced high migration inflow, a massive increase in real GDP and household disposable income combined with a decreasing unemployment rate. It seems that in Berlin, this favorable economic development resulted in higher demand for larger and more expensive dwellings, hence led to a strong liquidity decline at the upper end of the distribution. This again highlights the importance of the CQR approach.

**Figure 3.3** Estimated quantile regression coefficients of the quarterly dummy variables



**Notes:** The figure displays the development of the coefficients  $\beta_k^t$  of the quarterly dummy covariate for the 0.2-, 0.5- and 0.8-quantile for each of the seven cities  $k$  over time. Altogether, the data consists of 482,196 observations on the residential rental market. The sample period is 2013 Q1 to 2017 Q4.

### 3.5.2.2. Housing characteristics

It is possible to cluster the housing characteristics into three groups based on their impact on time on market. Unsurprisingly, asking rent has a consistently positive impact on time on market, suggesting that an increase in rent increases the time a dwelling is advertised on the market. The coefficients of first occupancy and renovation (besides the statistically insignificant 0.8 quantile in Cologne), as well as the equipment variable with bathtub, also show positive effects. Living area (besides the statistically insignificant 0.7 quantile in Dusseldorf) as well as number of rooms show the opposite impact for all quantiles across the seven cities. For the other dwelling characteristics, the coefficients yield distinct effects, depending on the respective quantile and location of the dwelling. Coefficients of variables of interest, supporting the hypotheses stated in Section 3.2, are shown in Figure 3.4.

As mentioned above, the covariate of asking rent is positive and statistically significant for each quantile of all cities. The magnitude of impact on time on market decreases with increasing time on market for all cities. This might be due to the fact that the more illiquid quantiles are on average characterized by larger living areas, more rooms and are further from the city center. With the population in the seven metropolises consisting to a great extent of young single households, larger dwellings are less in demand, so that the reaction of time on market to a rental price increase is less pronounced. Furthermore, higher time on market quantiles are also characterized by higher asking rents. Hence, an additional reason might be that a further rental increase starting from a relatively high rental level does not affect demand and liquidity that strong. The impact of an increase in asking rent is weakest for dwellings located in Dusseldorf. A reason for that might be that the overall socioeconomic and economic development has been relatively weak in Dusseldorf compared to the other cities. According to data provided from Oxford Economics, Dusseldorf experienced a relatively low increase in population, working population and household disposable income during the observation period. Though the vacancy rate has been decreasing during the last years, vacancy is still high in comparison to the other metropolises. Additionally, building completion has been increasing and is at a relatively high level, while the city exhibits the second largest purchasing power per household in this sample. These factors influence real estate prices and rents. Consequently, asking rents in Dusseldorf have been rising below-average relative to the other considered cities. Hence, these socioeconomic and economic factors dampen the impact on time on market. The strongest impact for the more liquid quantiles is found for Stuttgart and Munich and switches to Berlin for the 0.8-quantile. This is not surprising, as rents in Munich as well as Stuttgart are at a very high level compared to the other cities. Accordingly, further rental price increases substantially affect time on market. The effect in Berlin, however, is considerably weaker for all quantiles and only exceeds the impact in Munich and Stuttgart for the most illiquid dwellings. While a ten percent increase in asking rent within the most liquid quantile results in an

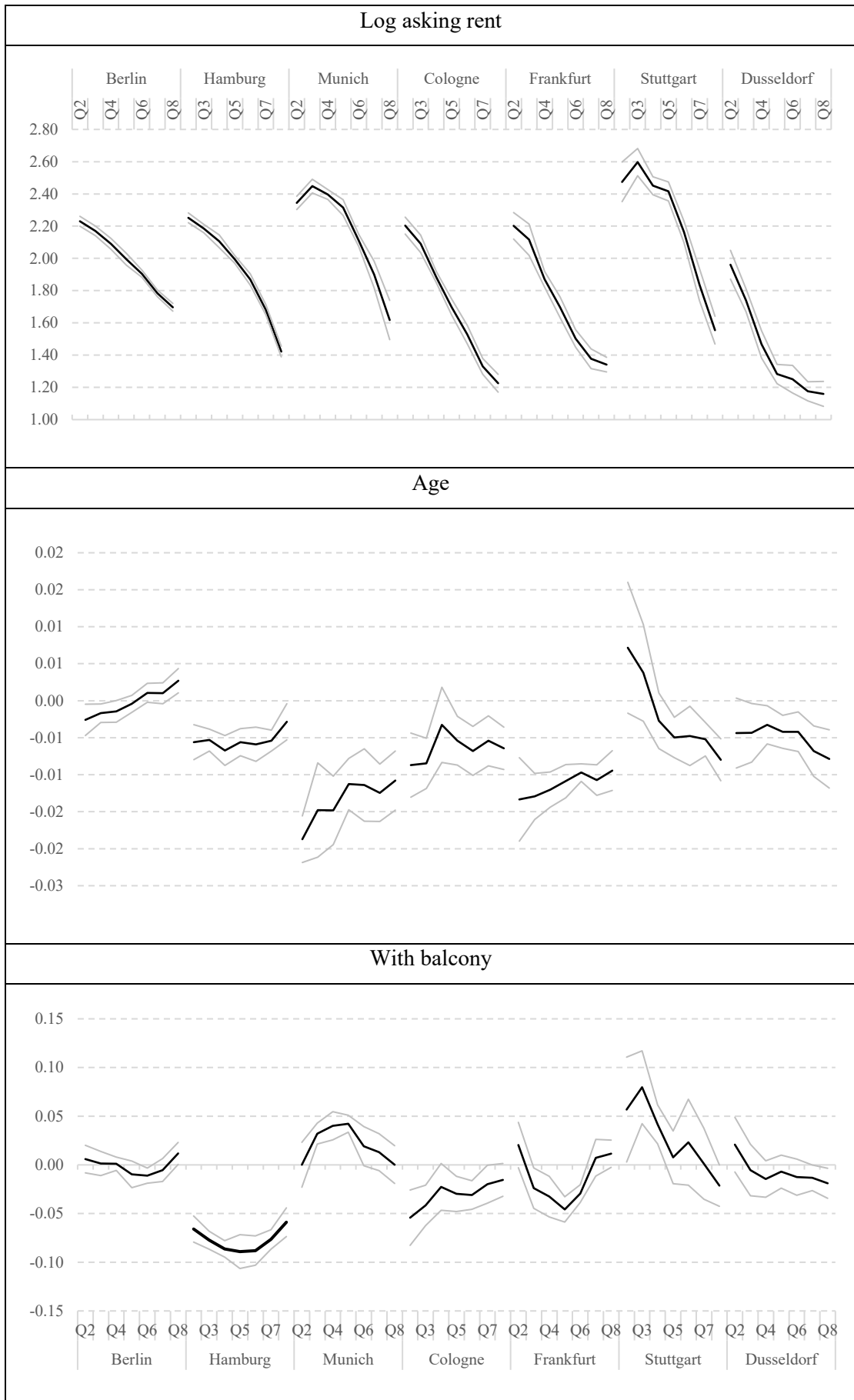
8.5% to 10.8% higher time on market in Dusseldorf and Stuttgart, the impact decreases to between 4.9% and 7.3% higher time on market in Dusseldorf and Berlin for the most illiquid quantile.

The impact of an increase in living area as well as the number of rooms shows the expected opposite pattern, as more space and a higher usability are positive factors for the marketability of a dwelling, all else equal. With growing illiquidity, the effect of living area, as well as of the number of rooms on the marketing time diminishes. For all quantiles, the strongest impact of an increase in living area is found for dwellings in Stuttgart. However, the impact of the number of rooms on time on market is lowest in Stuttgart. Thus, relative to the other cities, people in Stuttgart prefer larger apartments with less rooms. However, in Hamburg the importance of an additional room is highest among the seven cities, whereas the effect of the living area is lowest for highly liquid dwellings. A similar picture holds true for Berlin. These findings emphasize the presence of heterogeneity between the cities and the importance of analyzing each city individually.

In contrast to the uniform direction of impact of asking rent, the effect of a change in age does not have a consistent effect on time on market across all quantiles and cities, nor is it statistically significant for all of them. All statistically significant coefficients are negative, hence an increase in age reduces time on market, meaning that older dwellings are rented out faster. However, regarding the statistically insignificant coefficients as well, exhibits that within Berlin and Stuttgart the direction of effect changes along the time on market distribution. While in Berlin an increase in age leads to a longer time on market at the upper tail of the distribution, in Stuttgart this direction of effect is true at the lower tail. Consequently, the proportional hazards assumption underlying the Cox PHM cannot be verified for the covariate of age for all cities. This supports the CQR approach, where the effect of a covariate is analyzed for each quantile individually.

The direction of impact a change in certain equipment characteristics has on time on market does not only differ between the cities but also between the time on market quantiles within a city. For example, a balcony negatively influences time on market in Hamburg but positively in Munich in each quantile. In Berlin, Frankfurt and Dusseldorf the impact of a balcony changes its sign along the time on market quantiles, giving rise to the need of censored quantile regressions. The effect of an elevator on time on market is positive in Berlin and negative in Munich. Moreover, in Frankfurt the effect is significantly negative at the 0.2 quantile, however, significantly positive at the 0.6, 0.7, and 0.8 quantile. A similar picture is visible for the effect of a built-in kitchen. The impact on time on market is negative in Berlin, Munich, Cologne, Frankfurt, and Dusseldorf but positive in Hamburg for each quantile. In Stuttgart in contrast the impact of a built-in kitchen is significantly positive at the 0.2 quantile and significantly negative at the 0.6, 0.7, and 0.8 quantile. These binary variables strongly show the need for regarding each city separately as well as the importance of the CQR approach.

**Figure 3.4** Estimated quantile regression coefficients of selected housing variables



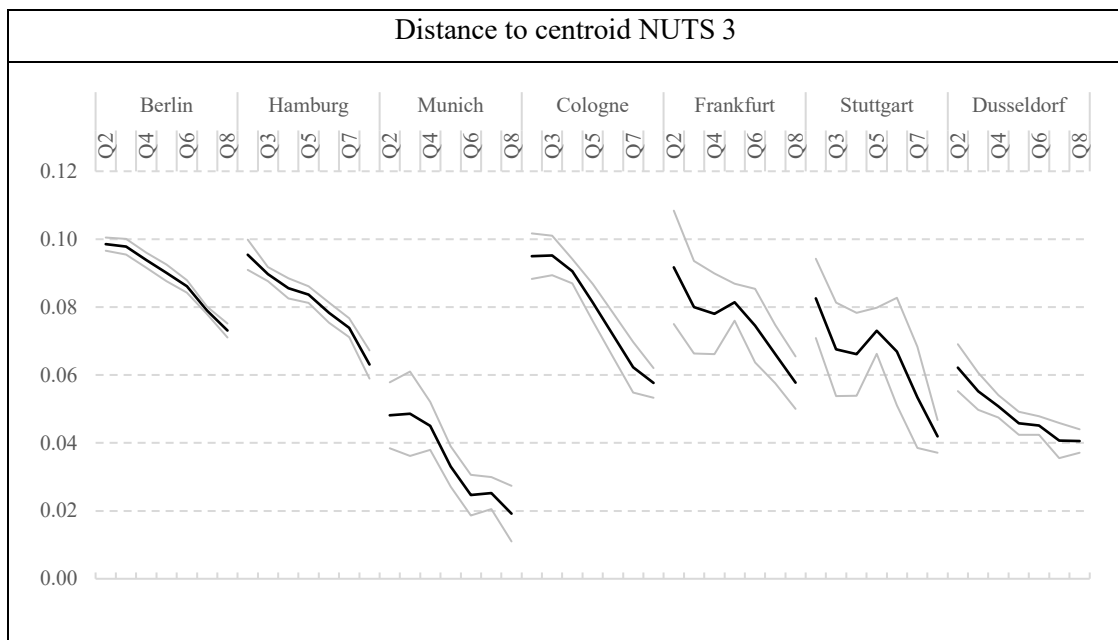
**Notes:** The figure displays the development of the coefficients  $\beta_k^\tau$  of the housing covariates across the liquidity quantiles  $\tau$  for each of the seven cities and the respective confidence intervals. The impact of an individual coefficient is insignificant if the confidence interval includes zero. The data consists of 482,196 observations on the residential rental market. The sample period is 2013 Q1 to 2017 Q4.

### 3.5.2.3. Spatial gravity variables

Since the study analyzes the seven largest German cities geographically each restricted by its NUTS 3 boundaries, the distance of a dwelling to the NUTS 3 center is of particular interest as it can be interpreted as its distance to the geographical city center.

Figure 3.5 shows that the effect of the distance to the geographical city center on time on market is positive and statistically significant for each city and each quantile. Hence, a higher distance to the center extends the time a rental dwelling is listed on the market. The effect is strongest for the highly liquid quantiles and becomes weaker for more illiquid ones. A possible explanation might be the fact that the most liquid dwellings are located in the city centers, thus moving away from the center has the strongest effect. In contrast, the more illiquid dwellings are located at the outskirts where one kilometer closer to or further from the center makes little difference. Across the whole distribution, the effect is highest in Berlin and lowest in Munich. With on average 8.44 kilometers, the rental dwellings in Berlin display on average the highest distance to the approximated city center. People in sprawling Berlin evidently have higher preferences for living closer to the city center compared to the other cities, especially Munich. A reason for this might be the location of popular residential areas right across the metropolitan area of Berlin, as well as the still relatively low asking rents in the city center compared to other cities. The distance to the ZIP code center is of less interest and has no explanatory power.

**Figure 3.5** Estimated quantile regression coefficients of the spatial gravity variable



**Notes:** The figure displays the development of the coefficients  $\beta_k^\tau$  of the spatial gravity covariate distance to the NUTS 3 center across the liquidity quantiles  $\tau$  for each of the seven cities and the respective confidence intervals. The impact of an individual coefficient is insignificant if the confidence interval includes zero. The data consists of 482,196 observations on the residential rental market. The sample period is 2013 Q1 to 2017 Q4.

### 3.5.2.4. Socioeconomic characteristics

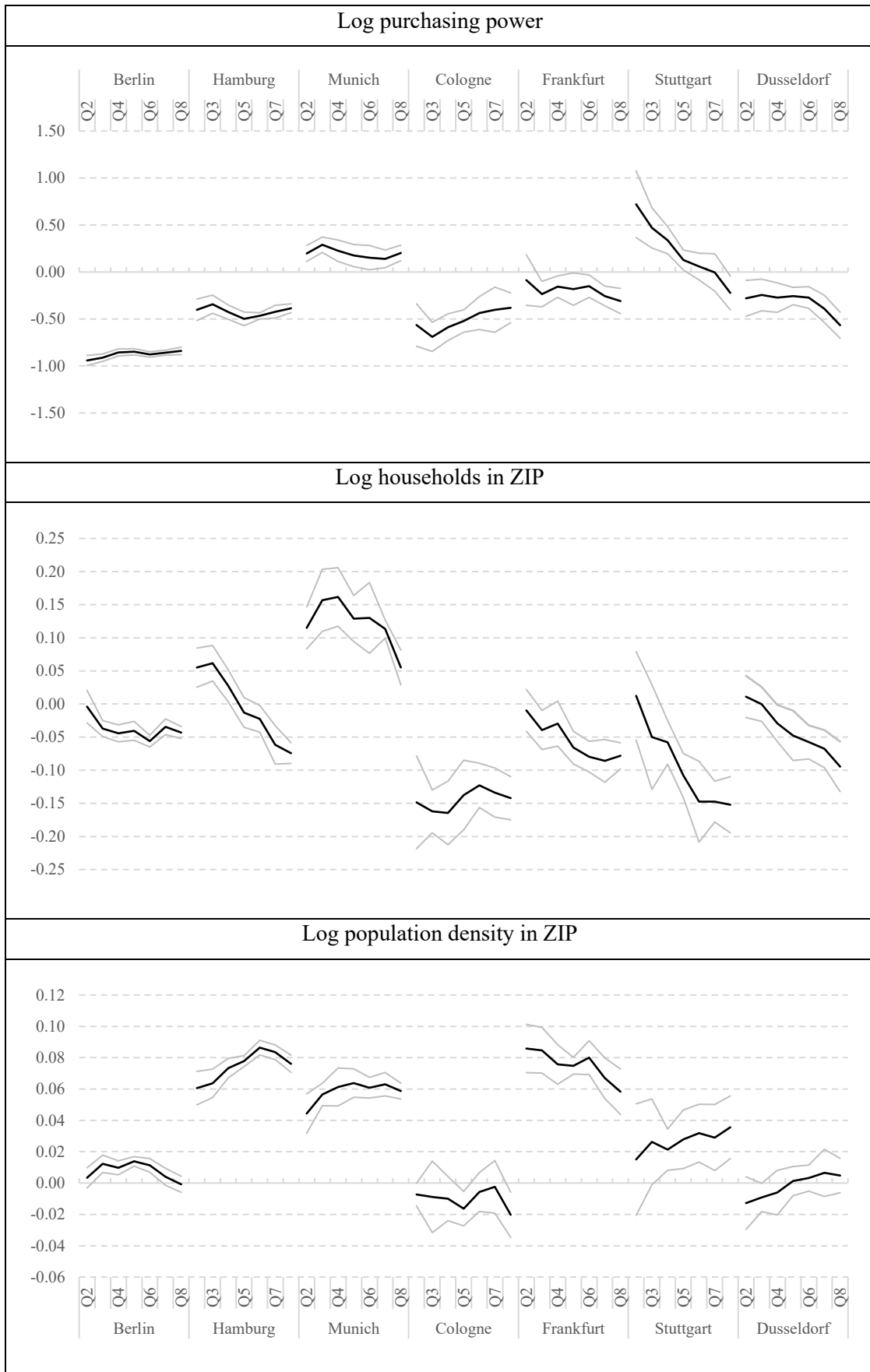
Each of the socioeconomic variables exhibits a quite different impact on time on market depending on the time on market quantile as well as the city.

An increase in purchasing power leads *ceteris paribus* to a shorter time on market in five of the seven cities. In these cities purchasing power drives the demand for residential real estate on the rental market and thus results in a lower marketing time. The impact is strongest in Berlin across all quantiles. Berlin shows by far the lowest average purchasing power. The substantial reduction in marketing time caused by an increase in purchasing power might indicate that people would be willing to spend more of the additional income on housing than in other cities. In Munich, however, the effect of purchasing power is time on market-prolonging. Hence, this points to the importance of considering the cities separately. Furthermore, the effect in Stuttgart exhibits a different sign depending on the respective quantile. Consequently, this supports the use of CQRs. Comparing the results of the CQR to the results of the Cox survival regressions, where the coefficient of purchasing power was only significant for three of the seven cities, reveals the huge heterogeneity within each city.

The effect of the number of households in a ZIP code area on time on market reveals a similar picture. While Berlin, Cologne, and Frankfurt show the expected demand effect, meaning that an additional household increases the demand for space and thus time on market decreases, in

Munich the impact is again time on market-prolonging. In the three remaining cities the impact on time on market even switches from positive to negative, implying a violation of the proportional hazards assumption. The direction of impact of population density is also not consistent for all time on market quantiles in two of the seven cities. These findings strongly emphasize to analyze the seven cities separately and to permit the direction of the effect to change with time on market quantiles and hence support the preference of the CQR approach over the Cox PHM.

**Figure 3.6** Estimated quantile regression coefficients of the socioeconomic variables



**Notes:** The figure displays the development of the coefficients  $\beta_k^\tau$  of the socioeconomic covariates across the liquidity quantiles  $\tau$  for each of the seven cities and the respective confidence intervals. The impact of an individual coefficient is insignificant if the confidence interval includes zero. The data consists of 482,196 observations on the residential rental market. The sample period is 2013 Q1 to 2017 Q4.

## Conclusion

Academic research on the factors affecting the liquidity of dwellings on the rental market is scarce and for most countries non-existent. Over the last years, the improvement in computational power and the ability to gather and store large amounts of data opens new research fields, especially when it comes to the application of advanced econometric methods. To the best of the authors' knowledge, this study is the first to apply a censored quantile regression (CQR) in order to explore the determinants of liquidity with an extensive dataset, comprising almost 500,000 observations across the seven largest German residential markets.

The main findings can be summarized as follows. For many covariates, consistent signs of the regression coefficients were found across the quantiles of the time on market distribution. However, for some covariates in individual cities, the impact of a change in the explanatory variables differs in direction between the liquidity quantiles. Hence, the proportionality assumption underlying the Cox proportional hazards model (PHM) is violated for those covariates in the associated cities. In addition, the impact of a change in the explanatory variables differs in magnitude and significance across the liquidity quantiles. In contrast to the traditional Cox proportional hazards regression, the CQR accurately detects these differences as it distinguishes between differential effects across quantiles. Furthermore, it is evident that the magnitude, significance, and direction of the impact of the covariates on time on market is quite different across the cities. These findings emphasize the importance of considering each city individually for a more detailed analysis and understanding of the respective rental real estate market. To summarize, the CQR applied on each city individually provides a more complete statistical analysis than commonly used approaches. The results of the individual city might depend on its economic and socioeconomic development during the last years. This overall development might include factors like population, working population, real GDP, household disposable income, unemployment rate, vacancy rate and construction activity. Additionally, the socio-demographic structure of the population regarding the age distribution, the distribution of the number of household members and the income distribution, among others, might play an important role.

With this detailed market assessment, landlords of dwellings should be able to infer how fast they will be able to let them, or which actions to take in order to increase their marketability. The variation in the impact of individual covariates on time on market across the liquidity quantiles and across the cities reveals the very distinct market characteristics in terms of marketability and

location. While the study uses the rental market of the seven largest German cities, it is of course possible to adapt the methodology to more cities or conurbations and other international real estate markets in order to examine the individual time on market quantiles.

As addressed in Section 3.4 Data and Descriptive Statistics the covariate asking rent is defined as the rent initially set by the landlord. Rent negotiations as well as contract rents are not available for the German residential rental market. On the residential rental market, the assumption that asking rents operate as a “take it or leave it option” to the tenant seems to be plausible as rent negotiations, especially in the overheated markets of the seven largest cities considered in this paper, are rather an exception. Nevertheless, the availability of contract rents would further improve the findings of this paper and moreover would expand the range of possible research questions. In general, using asking rents might come at the cost of a potential upward bias of rents. A lower contract rent than the initially asked rent might be possible. This scenario seems most likely for dwellings with a long marketing time. Consequently, in this paper the results in the time on market quantiles at the upper end of the distribution might be biased with respect to rents. The availability of contract rents would allow to estimate the impact of initial overpricing on time on market. Specifically, does initial mispricing significantly lead to longer marketing times? Furthermore, it would be interesting to investigate in which time on market quantiles the contract rent deviates from the asking rent. Another important covariate might be transaction costs. It could be investigated whether transaction costs (e.g. broker fees) lead to shorter or longer marketing times. It could be assumed, that landlords with higher transaction costs prefer to let a dwelling more quickly. That in turn might influence the initial price setting. Furthermore, incorporating further building characteristics, like the type of building (e.g. multi-family building, old building, detached building), the number of floors, and the floor level of the respective dwelling might alter the results. Especially, some of the binary housing attributes like with balcony or with elevator might yield more distinct results. Further research should also concentrate on a deeper analysis of an individual city. The city center might be specified not only geographically but rather in terms of an economic and sociodemographic city center. Furthermore, the proximity to local public transport and other amenities, tourism hotspots, commuting flows as well as the distance to adjacent urban areas should be included in the analysis. The type of neighborhood, school quality and noise, e.g. from a highway, should be considered as well. Including this mass of spatial variables, would lead to a more precise and detailed spatial mapping of a city. These variables would probably change the impact of the current variable “distance to the NUTS 3 center”. Moreover, the spatial variables are assumed to have a different effect on different dwelling and inhabitant types and thus in different time on market quantiles. Moreover, the impact of competition on the rental market, for example modelled as spatial concentration of simultaneously offered similar dwellings in a neighborhood, on time on market might be of further interest. The turnover of similar rental dwellings and the market tightness in a neighborhood might be an indicator of the bargaining power of landlords and tenants. As these competition variables are

enormously associated with the rental price setting, they might influence the current results of the variable asking rent. This might depend on the respective location and type of dwelling. Hence, these results might be different for different time on market quantiles. In addition, a counterfactual decomposition could reveal whether the impact on time on market is attributable to a pure change in the characteristics of the dwellings or a shift in the assessment of characteristics.

## References

- Allen MT, Rutherford RC, Thomson TA (2009) Residential asking rents and time on the market. *The Journal of Real Estate Finance and Economics* 38:351-365
- Anglin PM, Rutherford RC, Springer TM (2003) The trade-off between the selling price of residential properties and time-on-the-market: The impact of price setting. *The Journal of Real Estate Finance and Economics* 26:95–111
- Bassett Jr GW, Hsiu-Lang C (2002) Portfolio style: Return-based attribution using quantile regression. In: Fitzenberger B, Koenker R, Machado J (eds.) *Economic Applications of Quantile Regression, Studies in Empirical Economics, Physica, Heidelberg*, pp 293-305
- Belkin J, Hempel DJ, McLeavey DW (1976) An empirical study of time on market using multidimensional segmentation of housing markets. *Real Estate Economics* 4:57-75
- Benefield JD, Hardin WG (2015) Does time-on-market measurement matter? *The Journal of Real Estate Finance and Economics* 50:52-73
- Bentzien V, Rottke N, Zietz J (2012) Affordability and Germany's low homeownership rate. *International Journal of Housing Markets and Analysis* 5:289-312
- Beyerlein A, Toschke AM, von Kries R (2008) Breastfeeding and childhood obesity: shift of the entire BMI distribution or only the upper parts? *Obesity* 16:2730-2733
- Buchinsky M (1994) Changes in the US wage structure 1963-1987: Application of quantile regression. *Econometrica* 62:405-458
- Buchinsky M (1995) Quantile regression, Box-Cox transformation model, and the US wage structure, 1963–1987. *Journal of Econometrics* 65:109-154
- Cajias M, Freudenreich P (2018) Exploring the determinants of liquidity with big data - market heterogeneity in German markets. *Journal of Property Investment and Finance* 36:3-18
- Chaudhuri P, Doksum K, Samarov A (1997) On average derivative quantile regression. *The Annals of Statistics* 25:715-744
- Cox DR (1972) Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34:187-220
- Crowley J, Hu M (1977) Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* 72:27-36
- Federal Statistical Office (2018) Baufertigstellungen von Wohn- und Nichtwohngebäuden unter anderem nach Bauherren – Lange Reihen. [https://www.destatis.de/DE/Publikationen/Thematisch/Bauen/Bautaetigkeit/Wohnungsbau/BaufertigstellungenBauherrenPDF\\_5311201.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/Bauen/Bautaetigkeit/Wohnungsbau/BaufertigstellungenBauherrenPDF_5311201.pdf?__blob=publicationFile). Accessed 8 August 2018

Held T, Waltersbacher M (2015) Wohnungsmarktprognose 2030. BBSR-Analysen KOMPAKT, Nr. 7, Bonn. [https://www.bbsr.bund.de/BBSR/DE/Veroeffentlichungen/AnalysenKompakt/2015/DL\\_07\\_2015\\_.pd?\\_\\_blob=publicationFile&v=5](https://www.bbsr.bund.de/BBSR/DE/Veroeffentlichungen/AnalysenKompakt/2015/DL_07_2015_.pd?__blob=publicationFile&v=5). Accessed 28 February 2020

Hendricks W, Koenker R (1992) Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association* 87:58-68

Horowitz JL, Neumann GR (1987) Semiparametric estimation of employment duration models. *Econometric Reviews* 6:5-40

Kang H, Gardner M (1989) Selling price and marketing time in the residential real estate market. *Journal of Real Estate Research* 4:21-35

Koenker R, Bassett Jr GW (1978) Regression quantiles. *Econometrica* 46:33-50

Koenker R, Biliyas Y (2002) Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments. In: Fitzenberger B, Koenker R, Machado J (eds.) *Economic Applications of Quantile Regression, Studies in Empirical Economics*, Physica, Heidelberg, pp 199-220

Koenker R, Geling O (2001) Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association* 96:458-468

Lerbs O, Oberst A (2014) Explaining the spatial variation in homeownership rates: Results for German regions. *Regional Studies* 48:844-865

Liao WC, Wang X (2012) Hedonic house prices and spatial quantile regression. *Journal of Housing Economics* 21:16-27

Lin DY, Wei LJ (1989) The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* 84:1074-1078

Lüdemann E, Wilke RA, Zhang X (2006) Censored quantile regressions and the length of unemployment periods in West Germany. *Empirical Economics* 31:1003-1024

Machado J, Mata J (2000) Box-Cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics* 15:253-274

Miller NG (1978) Time on the market and the selling price. *Real Estate Economics* 6:164-174

Peng L, Huang Y (2008) Survival analysis with quantile regression models. *Journal of the American Statistical Association* 103:637-649

Portnoy S (2003) Censored regression quantiles. *Journal of the American Statistical Association* 98:1001-1012

Powell JL (1984) Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* 25:303-325

Powell JL (1986) Censored regression quantiles. *Journal of Econometrics* 32:143-155

Reisenbichler A (2016) A rocky path to homeownership: Why Germany eliminated large-scale subsidies for homeowners. *Cityscape* 18:283-290

Schmillen A, Möller J (2012) Distribution and determinants of lifetime unemployment. *Labour Economics* 19:33-47

Thomschke L (2015) Changes in the distribution of rental prices in Berlin. *Regional Science and Urban Economics* 51:88-100

Turnbull GK, Dombrow J (2006) Spatial competition and shopping externalities: Evidence from the housing market. *The Journal of Real Estate Finance and Economics* 32:391-408

Voigtländer M (2009) Why is the German homeownership rate so low? *Housing Studies* 24:335-372

Vonovia SE, CBRE GmbH (2016) *Wohnmarktreport Deutschland 2016*

Wehby GL, Murray JC, Castilla E, Lopez-Camelo JS, Ohsfeldt RL (2009) Prenatal care effectiveness and utilization in Brazil. *Health Policy and Planning* 24:175-188

Wood JH, Wood NL (1985) *Financial Markets*. Harcourt Brace Jovanovich, San Diego

Yavas A, Yang S (1995) The strategic role of listing price in marketing real estate: theory and evidence. *Real Estate Economics* 23:347-368

Zietz J, Zietz EN, Sirmans GS (2008) Determinants of house prices: a quantile regression approach. *The Journal of Real Estate Finance and Economics* 37:317-333

## 4 What are tenants demanding the most? A machine learning approach for the prediction of time on market

### Abstract

**Purpose** – This is the first article to apply a machine learning approach to the analysis of time on market on real estate markets.

**Design/methodology/approach** – The random survival forest approach is introduced to the real estate

market. The most important predictors of time on market are revealed and it is analyzed how the survival probability of residential rental apartments responds to these major characteristics.

**Findings** – Results show that price, living area, construction year, year of listing and the distances to the next hairdresser, bakery and city center have the greatest impact on the marketing time of residential apartments. The time on market for an apartment in Munich is lowest at a price of 750Vper month, an area of 60 m<sup>2</sup>, built in 1985 and is in a range of 200–400 meters from the important amenities.

**Practical implications** – The findings might be interesting for private and institutional investors to derive real estate investment decisions and implications for portfolio management strategies and ultimately to minimize cash-flow failure.

**Originality/value** – Although machine learning algorithms have been applied frequently on the real estate market for the analysis of prices, its application for examining time on market is completely novel. This is the first paper to apply a machine learning approach to survival analysis on the real estate market.

**Keywords** Residential, Housing, Time on market, Machine learning, Decision tree, Random forest

**Paper type** Research paper

**Acknowledgements** The authors especially thank PATRIZIA SE for contributing to this study. All statements of opinion are those of the authors and do not necessarily reflect the opinions of PATRIZIA SE or its associated companies.

## Introduction

The availability of large amounts of data yields great possibilities for highly precise predictions. One approach to deal with big data is machine learning. The mass of information combined with sophisticated statistical techniques leads to exciting possibilities for extremely precise and high-quality estimations. Towards the end of the twentieth century, machine learning techniques have found their way into the real estate market [1]. For the prediction of real estate prices, machine learning techniques have developed as an emerging alternative to the common hedonic models based on multiple regression. However, to the best of the authors' knowledge, there exists no application of machine learning techniques for the analysis of the time on market in the context of real estate market analysis. In this paper the time on market on the residential rental market of Munich is analyzed from a sample of 115,264 apartments based on micro data from the first quarter of 2013 to the fourth quarter of 2021.

The city of Munich is one of the economically most powerful and expanding metropolises in Germany and Europe. Munich is also characterized by high livability and living standard and hence accounts to the most attractive residential places. Moreover, the city of Munich is known as an economic, science and innovation hub in Germany. Hence, from 2013 to 2021 the city has been growing by around 6.6% to about 1.56 million inhabitants (Statistisches Amt der Landeshauptstadt München, 2022) due to strong migration and higher birth rates. This flourishing development, however, comes with the drawback of sparse supply, very low vacancies as well as increasing rents and prices on the residential market. Due to increasing prices and thus decreasing returns on residential rental apartments in Munich landlords are striving to rent out their apartments as quickly as possible. To estimate the time on market of residential rental apartments a machine learning approach based on the decision tree approach is applied in this article. According to Kok *et al.* (2017) there are great advantages of regression trees compared to multiple regression models. The assumptions about the distribution of data are more flexible, they can deal with linear and non-linear relationships, can process large amounts of data without consuming massive computing time, can handle categorical variables without dummy variables and the interpretation is simple and straightforward. As recommended by Breiman (2001), to redress the shortcomings of a single decision tree, an ensemble technique of decision trees, namely the random survival forest approach, is used in this paper. To the best of the authors' knowledge, this is the first paper to introduce the random survival forest approach to the real estate market. The findings might be of interest for current and future landlords to infer whether an apartment is highly demanded, to anticipate the respective marketing time and to take actions to decrease the respective marketing time. Furthermore, upon the findings in this article, private and institutional investors are able to derive real estate investment decisions and implications for portfolio management strategies and ultimately to minimize cash-flow failure. To precisely estimate the marketing time of an apartment, its location and with it its surrounding area are crucial. Čeh *et al.*

(2018) emphasize the importance of including geographic information to machine learning models for the prediction of real estate prices. Hence, the geographical coordinates of each apartment are used to calculate distances to several spatial variables to incorporate environmental and neighborhood information. These are included to the estimation of the marketing time of each apartment. Hence, the analysis allows us both, to determine the marketing time of an apartment more accurately and to detect the influence and importance of specific amenities on the time on market.

The study aims to answer the following questions regarding the residential rental real estate market.

1. What are the most important housing, social and spatial characteristics affecting the survival probability of residential rental dwellings in Munich?
2. How does the survival probability of residential rental dwellings respond to these major characteristics based on a random survival forest approach?

The remainder of this paper proceeds as follows: The next section presents a literature review, followed by a description of the underlying econometric model in Section 4.3. A detailed description of the dataset and the descriptive statistics can be found in Section 4.4. Estimation results are presented and discussed in Section 4.5. Finally, Section 4.6 concludes and addresses ideas for further research.

## Literature review

In recent years the application of machine learning techniques to the field of real estate research has gained great interest. Several articles in the field of real estate have been published applying different machine learning algorithms such as neural networks, fuzzy logic, decision trees, or ensemble methods. These methods have been applied especially to estimate real estate prices as well as to estimate the impact a predictor has on the property price. An early contribution to apply decision trees and neural networks on the real estate market has been made by Jaen (2002). In this study housing prices for 1,229 transactions in the city of Coral Gables (Florida) have been estimated and the most influential variables on housing prices have been identified. It concludes that the best results were received applying the classification and regression tree (CART) algorithm as it leads to the smallest mean absolute error (MAE) and uses the least number of covariates. The CART algorithm was also applied by Fan *et al.* (2006). They use a decision tree, that is a non-parametric approach, to overcome the potential problems of the widely used hedonic-based multiple regression approach. They determine the most important variables affecting housing prices and predict housing prices in the resale public housing market of Singapore. Özsoy and Sahin (2009) identify the most important determinants of housing prices in Istanbul (Turkey) also applying the CART approach. Pérez-Rave *et al.* (2019) apply a two-stage methodology under

a machine learning approach for inferential and predictive purposes on the housing market of Colombia. They first use the incremental sample with resampling (MINREM) methodology for the selection of the most important variables and then test the training and validation procedure for machine learning. They conclude that the applied algorithm results in stable models with less covariates. Kok *et al.* (2017) highlight the advantages of regression trees and also emphasize their limitations. To overcome these limitations Kok *et al.* (2017) recommend the application of tree ensemble methods like the random forest algorithm. One of the first studies to highlight the need of ensemble methods to improve prediction power was introduced by Breiman (2001). This paper describes the random forest algorithm, a machine learning algorithm which can be applied for classification as well as for regression problems. The random forest algorithm is an ensemble method and has been applied in numerous papers. Antipov and Pokryshevskaya (2012) apply the random forest method for mass appraisal on the residential real estate market of Saint-Petersburg (Russia) and empirically compare it to several other methods. They conclude that the random forest algorithm outperforms methods such as Chi-squared automatic interaction detection (CHAID), CART, k-nearest-neighbors (KNN), multiple regression analysis, artificial neural networks, and boosted trees. Yoo *et al.* (2012) also apply a random forest approach as well as Cubist and compare it to the traditional OLS regression for hedonic price models. Empirically applied to transaction data of the real estate market of Onondaga County, NY (USA) the random forest model leads to the highest model accuracy. Čeh *et al.* (2018) analyze the prediction of apartment prices in the city of Ljubljana (Slovenia) from 2008 to 2013. Therefore, they compare the random forest machine learning technique to common multiple regression models. They conclude that the predictive performance of the random forest method, measured by  $R^2$  values, sales ratios, the mean average percentage error (MAPE) and the coefficient of dispersion (COD), outperforms the commonly used hedonic models. These findings are confirmed by Nguyen (2018) who analyzed transaction data for five counties in the US and Dimopoulos *et al.* (2018) who estimate the transaction price of residential apartments in Nicosia (Cyprus). Both empirical studies find that random forest outperforms linear regression. Shahhosseini *et al.* (2020) empirically apply two popular housing datasets, Boston (USA) metropolitan housing and Ames (Iowa, USA) housing, to analyze different ensemble models for real estate price predictions. Comparing lasso regression, random forest, deep neural networks, XGBoost and support vector machines, they find XGBoost and random forest to outperform the other algorithms in terms of MAPE for the Boston metropolitan area dataset. For the Ames dataset they find lasso regression and random forest to perform best by means of MAPE. A recently published article by Gabrielli *et al.* (2023) analyzes the marginal effect of dwelling specific factors and neighborhood factors on the property price in Padova, Northern Italy, at three different time points: first, 2019 before the COVID-19 pandemic, second, 2021 two years after the first COVID-19 onset however before the war in Ukraine and third, 2023 one year after the war in Ukraine started. They apply a random forest feature importance analysis to exclude the variables with low significance levels from the

regression analysis. Combined with a multivariate regression, they find that the location of a building has the strongest impact on its price but also building characteristics play a major role. To conclude, the random forest algorithm is a widely applied machine learning technique and yields convincing results regarding price predictions in the field of real estate research.

All in all, current research concludes, that there exists no single technique that is superior to the others but emphasizes that tree ensemble techniques outperform other models. This paper follows the latest trends of literature on residential real estate pricing with machine learning techniques. However, instead of estimating the price of a residential apartment, the recent developments are transferred to analyzing the marketing time of an apartment. Hence, the target variable is the time an apartment is up for rent on the market, i.e. a duration is estimated. Therefore, another class of models is needed, namely survival models. The marketing time of a real estate property has been analyzed widely in previous years. The most established approach to measure time on market is the Cox (1972) proportional hazards model (PHM). Kluger and Miller (1990) were the first to apply the semi-parametric Cox PHM to the field of real estate research. Over the last years, this approach for the estimation of marketing time has been applied in several studies like Krainer (1999), Smith (2010), Hoerberichts *et al.* (2013) and Cirman *et al.* (2015), among others. Cajias and Freudenreich (2018) apply the Cox PHM to analyze the time on market of residential rental apartments in the seven largest German cities. They find that higher rental prices as well as larger living areas lead to longer marketing times in all seven cities. Cajias *et al.* (2020) set up a comparison of the Cox PHM to censored quantile regressions (CQR) to examine the time on market of residential rental dwellings in the seven largest German cities. The determinants of the time on market are examined for different time on market quantiles. Although the time on market of residential dwellings has been in the focus of a wide range of research, to the best of the authors' knowledge, this is the first paper to apply a machine learning approach to survival data on the real estate market. However, machine learning algorithms for the analysis of survival data have been applied in several other fields of research. Specifically, the random survival forests algorithm is a very popular and frequently used algorithm due to its high precision and its non-parametric nature. It builds on the original random forest algorithm introduced by Breiman (2001) and it is implemented following the same principles. The random survival forests algorithm applied in this study was introduced by Ishwaran *et al.* (2008). It has been applied especially in clinical and medical research. Cetin *et al.* (2021) perform a "fair" comparison of the commonly used Cox regression to random survival forest on colorectal cancer survival data. They state that for a "fair" comparison performance has to be compared on equivalent sets, i.e. the error of the Cox regression model must also be calculated over an out-of-bag (OOB) ensemble. Conducting a "fair" comparison they find that random survival forests outperform Cox regression in terms of discrimination performance. This finding is also supported by Miao *et al.* (2015) for data on cardiovascular disease.

In this article the random survival forests algorithm is applied to analyze the dependent variable, time on market, defined as the elapse of time a rental dwelling is on the market, belonging to the class of right-censored survival data.

## Econometric approach

As stated above the aim of this paper is to detect the most important determinants as well as their influence on the time on market of residential rental apartments in the city of Munich. In the real estate literature, analyzing prices, several machine learning techniques have been applied. Ensemble methods like bagging, boosting and random forest have become quite popular. The random forest algorithm as an ensemble method uses regression trees as base learners and embeds randomization into the learning process. Randomization, to avoid correlation of trees and hence to reduce variance, is carried out in a two-step procedure. First, a tree is grown by a randomly drawn bootstrap sample of the data. Second, at each node of the tree a randomly selected subset of covariates is chosen as candidate variables for splitting. Hence, the random forest algorithm approximates rich classes of functions with high prediction performance while keeping generalization error low.

The random forest approach goes back to Breiman (2001) and can be used for regression and classification analysis. The random survival forest approach, introduced by Ishwaran *et al.* (2008), extends Breiman's (2001) approach to the analysis of right-censored survival data and hence is very well suited for the analysis underlying this paper. The random survival forest approach is able to automatically deal with the common restrictions and problems regarding survival data. Conceptually, the random survival forest method used for the present analysis goes back to Ishwaran *et al.* (2008). We re-describe the algorithm in detail as this is the first paper applying a random survival forest in real estate literature. The dependent variable is the right-censored time on market  $(T_i, \delta_i)$  where  $i=1,2,\dots,n$  is the index for rental dwellings,  $\delta_i$  is the censoring status with  $\delta_i = 1$  if a dwelling is taken off the market (experienced the event) and  $\delta_i = 0$  otherwise and  $T_i$  is the time to event, either till the dwelling is taken off the market  $\delta_i = 1$  or until the end of the observation period. The independent variable is  $X_i = (x_{1i}, x_{2i}, \dots, x_{ki} \dots)$ , where  $k=1,2,\dots,K$  is the index for covariates. The algorithm can be described as follows:

1. Several  $B$  bootstrap samples are drawn from the initial data set. Each bootstrap sample includes on average  $2/3$  of the original data, the excluded  $1/3$  of the data is called out-of-bag data (OOB data).
2. A survival tree is grown for each of the  $B$  bootstrap samples. To grow a random forest with right-censored survival data the survival time  $T_i$  as well as the censoring status  $\delta_i$  must be considered. At each node of the tree,  $p$  candidates are chosen as splitting variables. The splitting criterion must consider the survival time and the censoring status. The candidate variable that leads to the maximum survival difference between the daughter nodes, is selected to split the node. The performance of a split is measured by tree node impurity.
3. The tree is grown to its total size with each terminal node  $h \in \mathcal{Q}$  having a minimum of  $d_{min} > 0$  deaths or otherwise until the discriminative power of a split is below a certain

threshold. The dependent variable time on market and the respective censoring value in a terminal node  $h \in \mathcal{Q}$  can be written as  $(T_{1,h}, \delta_{1,h}), (T_{2,h}, \delta_{2,h}), \dots, (T_{n(h),h}, \delta_{n(h),h})$ .

4. The cumulative hazard function (CHF) for each dwelling  $i$  is calculated from the terminal node of each tree. The hazard function measures the risk of a dwelling leaving the market and can be described as the probability of a dwelling being taken off the market at a certain point in time given that it has not been taken off the market before. Applying the Nelson-Aalen estimator, the estimator of the CHF for a terminal node  $h$  can be described as

$$\hat{H}_h(t) = \sum_{t_{l,h} \leq t} \frac{d_{l,h}}{Y_{l,h}}.$$

$d_{l,h}$  describes the number of dwellings who have experienced the event and  $Y_{l,h}$  are the dwellings at risk at time  $t_{l,h}$ . The CHF estimate for a terminal node defines the CHF for the respective tree according to  $H(t|x_i) = \hat{H}_h(t)$ , if  $x_i \in h$ , with  $x_i$  being the covariate of dwelling  $i$ . The ensemble CHF is received by averaging over the CHFs of the  $B$  survival trees

$$H_e^*(t|x_i) = \frac{1}{B} \sum_{b=1}^B H_b^*(t|x_i).$$

5. The prediction error for the ensemble CHF is estimated with OOB data using Harrell's concordance index (Harrell *et al.*, 1982). The C-index is defined as the percentage of concordant pairs. A C-index of 1 represents the best model prediction, whereas a C-index of 0.5 indicates a random prediction.
6. This procedure is repeated  $n_{rep}$  times, so that the overall error can be calculated as an average from several  $n_{rep}$  errors calculated previously. Thus, each iteration adds to the overall performance.

The estimation results can be interpreted using partial dependence (PD) plots. These exhibit the predicted conditional mean of the dependent variable, time on market, as a function of each covariate. The plots are based on partial dependence functions which exhibit the effect of a single covariate on the dependent variable given the average effects of all other variables. PD plots reveal insightful information, e.g., whether the relationship can be explained linearly or in a more complex manner. The plots are shown in Section 4.5.2 for the ten most important variables influencing the time on market, however, estimation includes all 63 independent variables. The importance of a specific covariate for the estimation of the dependent variable, time on market, is estimated with a technique called variable importance (VIMP). Therefore, the model is estimated with and without the specific variable and compared afterwards based on the respective performance. Hence VIMP of a covariate  $x$  describes the development of the misclassification error if the covariate  $x$  wasn't included in the estimation process. Large positive values of VIMP

suggest variables with strong predictive power. In contrast zero and negative values indicate variables with poor predictive power.

## Data and descriptive statistics

The estimation sample is composed of three merged datasets, containing information from 115,264 observations of single- and multi-family rental dwellings in the city of Munich over a period of nine years from the first quarter of 2013 to the fourth quarter of 2021. Observations regarding student housing, affordable housing, temporary housing, and retirement housing have not been incorporated into the dataset. Information on the rental dwellings was provided from Value Marktdaten which collects georeferenced real estate data from more than 100 German Multiple Listing Systems (MLS) such as ImmoScout, Immonet or Immowelt but also regionally focused marketplaces and newspapers for the whole German market. As the market leader of real estate data for Germany, Value Marktdaten has an own proprietary algorithm that identifies duplicates and harmonizes the sample. The variable of major interest, time on market, is defined as the number of weeks a dwelling was listed in the MLS calculated by the start and end date according to Benefield and Hardin (2015) for example. The asking rent is defined in absolute terms measured in euros per month. The common issue of reverse causality between time on market and rent, discussed by Yavas and Yang (1995) among others, is not addressed in this paper. The asking rent is set by the landlord at the beginning of the data generating process (DGP). The time on market is established as the time the dwelling is offered in the MLS in weeks given the initially set asking rent. The asking rent operates as a “take it or leave it option” to the tenant and thus rent negotiations are not considered. On the residential rental market of Munich this assumption is plausible as negotiations about the monthly rental payments are rather an exception. This is also suggested by Gröbel (2019) for asking data on the German housing market.

Gao *et al.* (2022) state that the factors influencing the real estate price, applied to this analysis the time on market, can be divided into two main categories: non-geographical features, i.e. specific dwelling characteristics, and geographical features, describing the surrounding environment. Following Mora-Garcia *et al.* (2019) dwelling characteristics can be separated into single dwelling and building characteristics. The first sub-group maintains the features of each specific dwelling located in a building. In this study the single dwelling characteristics are characterized according to the rental price, the living area, the number of rooms, the respective floor, and the binary variables of whether it contains a bathtub, a balcony, or a terrace. The latter sub-group contains the variables describing the peculiarities referring to the whole building, including the construction year of the building and the existence of an elevator. Hence, this study includes nine independent variables depicting the dwelling features, listed in Appendix Table 4.3.

What are tenants demanding the most? A machine learning approach for the prediction of time on market

Moreover, several geographical factors representing the location of a dwelling are included. Since the data is georeferenced, the surrounding area of each dwelling can be characterized precisely. The coordinates for the location variables stem from OpenStreetMap. Therefore, spatial gravity indicators, measuring the Euclidian distance of each dwelling to the surrounding neighborhood characteristics are included. In literature there exists no uniform classification of neighborhood characteristics. According to Ahlfeldt and Wendland (2017) neighborhood characteristics can be classified into four groups: natural amenities, consumption amenities, accessibility attributes and sociodemographic information. In this paper we summarize the first two dimensions to urban amenities. Following Gabrielli *et al.* (2023) urban amenities can be further classified into supply services or open-air amenities. Supply services include healthcare facilities (e.g. hospital, doctor, pharmacy and dentist), childcare and teaching facilities (e.g. kindergarten, school and university), commercial facilities (e.g. restaurants, bakery, supermarket, mall and hairdresser), sports facilities (e.g. sport center) and financial facilities (e.g. atm, post office). Open-air amenities are defined as public amenities like parks, playgrounds or public swimming pools. Furthermore, accessibility attributes include the distance of a dwelling to the city center as well as proximity to public transportation, such as bus stop, tram stop or railway station among others. Moreover, sociodemographic information, i.e. the purchasing power per household and the number of households at the ZIP code level, obtained from “Growth from Knowledge” (GfK), is included. By means of all these geographic variables the location of each dwelling can be characterized precisely. Table 4.1 shows the descriptive statistics for the variables of interest for the city of Munich. The mean time on market is 3.41 weeks. The median time on market with only 1.5 weeks is even lower, stating that half of the dwellings are on the market for maximal one-and-a-half-weeks. On average a dwelling in Munich costs 1178.19 € per month, is 70.37 m<sup>2</sup> divided into 2.4 rooms, located on the 2nd floor and constructed in 1976. The next hairdresser is on average located in 310 meters, bakery, and school in 250 and 350 meters respectively, is located 4.50 kilometers from the city center and the next windmill is 52.92 kilometers away. A ZIP code area is on average populated by 20,814.82 people divided into 11,453.13 households, with a mean purchasing power per household of 55,944.97 € per year.

**Table 4.1** Descriptive statistics

	<b>Mean</b>	<b>Median</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
Time on market	3.41	1.50	5.06	0.10	35.60
Price	1178.19	1070.00	552.52	120.00	3850.00
Living area	70.37	66.66	29.41	21.10	194.00
Year of construction	1976.30	1976.00	33.55	1900.00	2022.00
Year of listing	2017.50	2018.00	2.61	2013.00	2021.00
Distance hairdresser	0.31	0.25	0.26	0.00	2.12
Distance bakery	0.25	0.21	0.18	0.00	1.62
Distance city center	4.50	4.49	2.01	0.06	9.86
Distance school	0.35	0.32	0.20	0.00	1.11
Distance windmill	52.92	53.42	2.38	44.45	57.64
Month of listing	6.36	6.00	3.44	1.00	12.00
Number of rooms	2.40	2.00	0.97	1.00	8.00
Floor	2.11	2.00	1.72	-1.00	9.00
Purchasing power per household ZIP	55,944.97	55,415.67	6128.56	39,901.99	71,765.44
Number of households ZIP	11,453,13	12,007.00	3240.24	1705.00	17,021.00
Population ZIP	20,814.82	22,665.00	6103.57	3106.00	31,271.00

**Note(s):** N = 115,264

**Source(s):** Table created by authors

## Results

### 4.1.1 Variable importance

In a first step the importance of each covariate for the dependent variable, time on market, is determined. By far the strongest influence on the marketing time of a dwelling has the price of the dwelling. This is not surprising as the price is the most remarkable criterion of a property and of major interest. In addition, the living area and the construction year of the apartment have a great impact. The time dummy variable that describes the development of the time on market over the years from 2013 to 2021 provide further insights. The next category of influential variables are distance variables. The distance of a dwelling to the next hairdresser, the next bakery, the city center, and the nearest school impact the time on market. This is interesting as these variables can be viewed as representatives of the categories, distances to social amenities, businesses of the daily need, and teaching facilities. These proximities obviously are of great importance for the attractiveness and hence the marketing time of a dwelling. Additionally, the distance to the closest windmill affects the time on market. This proximity is connoted rather negatively. Furthermore, the seasonal aspect of renting out an apartment affects the time on market. The availability of a bathtub, an elevator, a balcony, or a terrace apparently have no influence on the marketing time

as their variable importance is 0. This suggests that these variables are perfectly priced and have no further impact beyond the price. Also surprising is the fact that no covariate depicting the distance to the next public transport is among the ten most influential variables. The variable importance as well as the relative importance of the ten most influential variables on the time on market are depicted in Table 4.2.

**Table 4.2** Variable importance of 10 most influential variables on predicted survival time

	<b>Importance</b>	<b>Relative Imp</b>
1. Price	0.0819	1.0000
2. Living area	0.0320	0.3907
3. Year of construction	0.0229	0.2792
4. Year of listing	0.0138	0.1690
5. Distance hairdresser	0.0055	0.0670
6. Distance bakery	0.0053	0.0652
7. Distance city center	0.0048	0.0583
8. Distance school	0.0044	0.0542
9. Distance windmill	0.0044	0.0533
10. Month of listing	0.0043	0.0526

Source(s): Table created by authors

#### 4.1.2 Results of the random survival forest model

In a second step the impact of the ten most influential predictors on the time on market out of the 62 covariates is depicted in Figure 4.1 and described below. The interpretation of the relation between the respective covariate and the time on market is given ranked according to the variable importance.

1. Price: In general, the higher the price, the longer the dwelling is on the market. For relatively low-priced apartments from 120 € to about 750 € the time on market is decreasing with increasing prices. The marketing time is lowest for prices around 750 €. Hence, the cheapest apartments are not the most demanded. A constant increase of marketing time with prices is observed for all apartments from 900 € onwards. The increase in marketing time with increasing prices is especially strong from 1300 € to 2000 €. Beyond that price less and less apartments are listed.
2. Living area: For small apartments ranging from 21 m<sup>2</sup> to about 60 m<sup>2</sup> the marketing time is lower the larger the living area with a minimal marketing time at about 60 m<sup>2</sup>. This is not surprising as this size range is probably the most demanded as most households in Munich are single households (53.9%) or two person households (24.9%) (Statistisches Amt der Landeshauptstadt München, 2022). From that point onwards each additional m<sup>2</sup>

increases the time an apartment is on the market, depicting the decreasing demand for larger apartments.

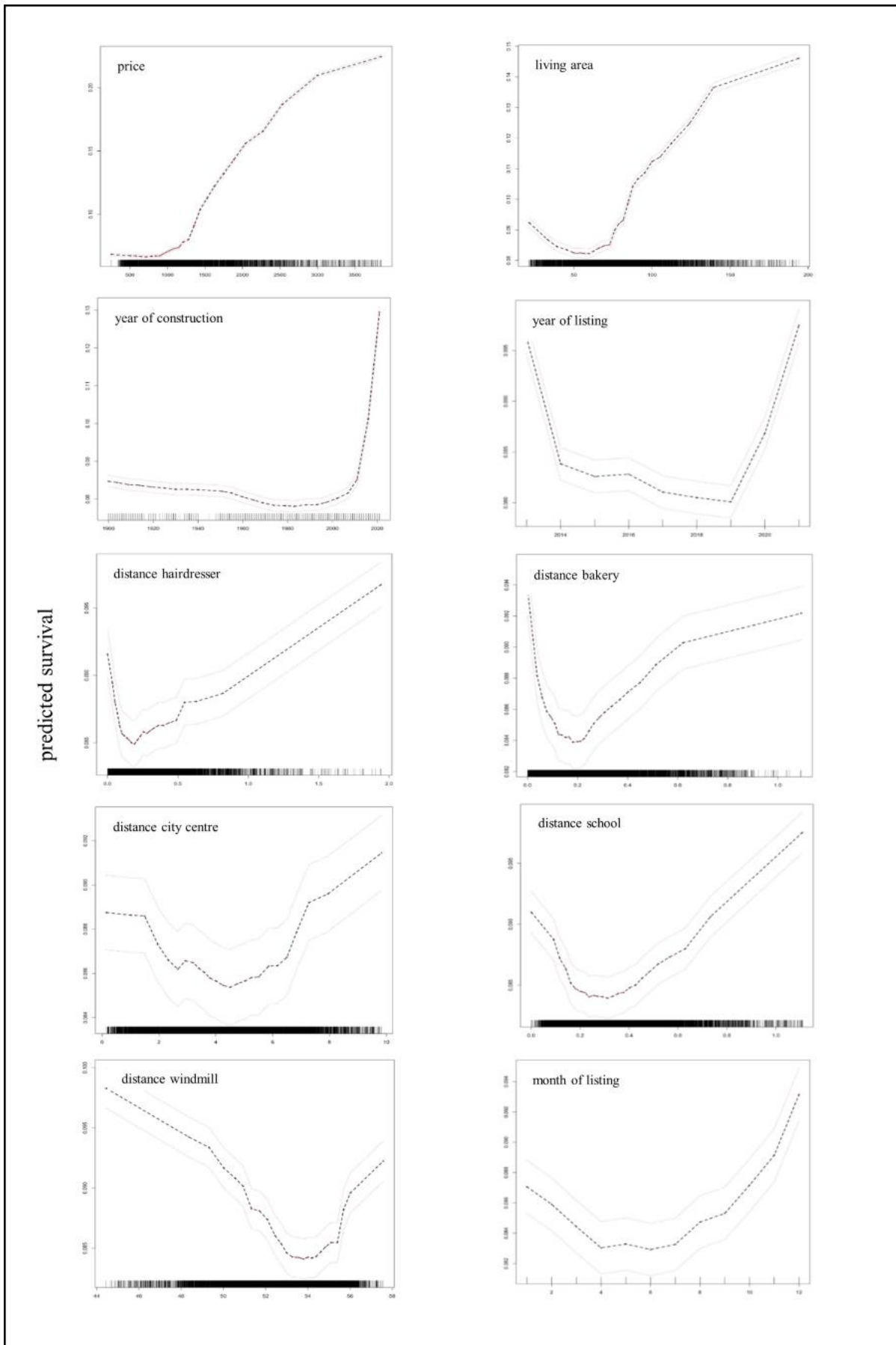
3. Year of construction: The marketing time has been decreasing for apartments built from 1900 till 1985, with the lowest marketing time for apartments in buildings built around 1985. Apartments built in the recent past exhibit the longest marketing time. The rise in marketing time is especially strong for apartments built from 2007 till today. Hence, newly constructed buildings face the longest marketing times.
4. Year of listing: This variable represents the change in the marketing time over the observation period from 2013 to 2021. Generally, the marketing time has been decreasing from 2013 to 2019, a period of flourishing and prosperous economic conditions. Recently, from 2019 to 2021 time on market has been sharply increasing, with this period being characterized by the Corona pandemics and hence economic uncertainties.
5. Distance hairdresser: The distance to the closest hairdresser is the most important distance variable and can be viewed as a representative for the distance to social amenities. An apartment directly next to a hairdresser exhibits a relatively high marketing time which decreases with every meter until a minimum time on market is reached at about 200 meters. From then on, every meter prolongs the time an apartment is listed for rent on the market.
6. Distance bakery: A very similar pattern is visible regarding the distance to the nearest bakery. Living in direct proximity to a bakery exhibits the highest time on market among all bakery distances. There is a sharp decline in marketing time for each meter moving away from the bakery until a distance of 200 meters, where the lowest time on market can be achieved. From then on, the farther an apartment is away from a bakery the longer the apartment stays on the market.
7. Distance city center: The distance to the city center has a U-shaped impact on the time on market. In a range of up to 4.5 kilometers from the city center the time on market decreases the farther the apartment is from the center. The minimum marketing time can be reached at about 4.5 kilometers from the center. From then on, the farther you move from the city center, the longer an apartment stays on the market. Hence, people prefer not to live directly in the city center but in close proximity.
8. Distance school: The picture here is very similar to the bakery distance. Apartments right next to a school show relatively high marketing times, probably caused by the noise and traffic around schools. Moving up to 300 meters marketing time is decreasing. From that point on, apartments stay longer on the market the higher the distance to the school.
9. Windmill distance: Living near a windmill certainly shows the highest marketing time. The farther you move from a windmill the more attractive an apartment becomes, until a distance of about 54 kilometers, where the minimal time on market can be achieved. From

then on, the marketing time increases again, however, the maximum distance of an apartment to a windmill is already reached at 57.64 kilometers.

10. Month of listing: The month an apartment was listed on the market represents the seasonality aspect. While apartments listed on the market in April or June exhibit the lowest time on market on average, the marketing time is longest in November and December. Hence, time on market is lower for apartments listed during spring and summer than during fall and winter. This implies that people rather move during warm months, and it takes longer to rent out an apartment during the cold season.

To summarize, the time on market for an apartment in Munich is lowest on average at a price of 750 € per month, an area of 60 m<sup>2</sup> and was built in 1985. Furthermore, the next hairdresser as well as bakery are 200 meters away, the nearest school is in a distance of 300 meters, the next windmill is about 54 kilometers and to the city center it is about 4.5 kilometers. Moreover, the time on market is lowest for apartments listed on the market in April or June. Generally, in an annual comparison, apartments listed in 2019 were the fastest to be rented out.

**Figure 4.1** Partial dependency plots for the ten most influential variables on predicted survival time



## **Conclusion**

In recent years the application of machine learning techniques has become quite popular in real estate research. For the prediction of real estate prices, machine learning techniques have developed as a fruitful alternative to the common hedonic models based on multiple regression. In contrast, the second insightful component on the residential real estate market, namely the time on market, has not yet been researched applying a machine learning algorithm. To the best of the authors' knowledge, this is the first paper to apply a machine learning approach to survival data on the real estate market. A random survival forest approach was introduced. The aim of this study was, first, to reveal the most important predictors of time on market and analyze how the survival probability of residential rental apartments responds to these major characteristics. The underlying sample consisted of micro data of 115,264 single- and multi-family rental apartments in the city of Munich over a period of nine years from the first quarter of 2013 to the fourth quarter of 2021. By far the strongest influence on the marketing time of an apartment had the price of the apartment, followed by the living area, the construction year and the year of listing. The next category of influential variables are distance variables. The distance of a dwelling to the next hairdresser, the next bakery, the city center, the nearest school, and the next windmill impact the time on market. Furthermore, the seasonal aspect of renting out an apartment affected the time on market.

The estimation results were interpreted using partial dependence (PD) plots that visualize the relationship the machine learning algorithm detects between the dependent variable time on market and its determining factors in the underlying dataset. The insights are interesting for further research as well as for practical implications. Real estate managers can derive their strategies regarding their portfolio management and can adjust their portfolio according to the predicted survival probabilities. With a profound understanding of the time on market, its determinants and its behavior to changes in the determinants, real estate professionals as well as private investors are able to minimize their cash-flow failure. With this detailed time on market assessment, current and future landlords of apartments in Munich should be able to infer how long it will take to let their apartments or which actions to take to advance the letting process. From a statistical perspective, the random survival forest algorithm bears several advantages compared to commonly applied regression models. It is absolutely data driven and hence does not depend on any model assumptions. It establishes a model that optimally explains the underlying data. The assumptions about the distribution of data are more flexible and they can deal with linear and non-linear relationships. It is able to deal with high dimensional data without massive computing time and overcomes limitations of common regression models, like overfitting or convergence issues. Further research on the demand factor time on market can build upon this article and is surely needed.

As the findings are limited to the city of Munich, an extension to other markets could be a fruitful starting point for future research. It might be interesting whether the most influential variables on the time on market vary between different markets. Furthermore, a comparison between the city of Munich and its outskirts might be an interesting indication of differences between urban and rural residential demand. Following Gabrielli *et al.* (2023) a comparison at different time points could provide further insights into the reaction of time on market to changes in its dwelling specific and neighborhood specific factors. Furthermore, a fair model comparison to the commonly applied Cox proportional hazard model could be insightful.

## References

Ahlfeldt, G.M. and Wendland, N. (2017), “The micro-cosmos of German cities: new insights into the supply of and demand for urban amenities”, Just, T. and Maennig, W. (Eds), *Understanding German Real Estate Markets*, Springer, pp. 79-88.

Antipov, E.A. and Pokryshevskaya, E.B. (2012), “*Mass appraisal of residential apartments: an application of random forest for valuation and a CART-based approach for model diagnostics*”, *Expert Systems with Applications*, Vol. 39 No. 2, pp. 1772-1778, doi: 10.1016/j.eswa.2011.08.077.

Benefield, J.D. and Hardin, W.G. (2015), “*Does time-on-market measurement matter?*”, *The Journal of Real Estate Finance and Economics*, Vol. 50 No. 1, pp. 52-73, doi: 10.1007/s11146-013-9450-z.

Borst, R.A. (1995), “*Artificial neural networks in mass appraisal*”, *Journal of Property Tax Assessment and Administration*, Vol. 1 No. 2, pp. 5-15.

Breiman, L. (2001), “*Random forests*”, *Machine Learning*, Vol. 45 No. 1, pp. 5-32, doi: 10.1023/A:1010933404324.

Cajias, M. and Freudenreich, P. (2018), “*Exploring the determinants of liquidity with big data—market heterogeneity in German markets*”, *Journal of Property Investment and Finance*, Vol. 36 No. 1, pp. 3-18, doi: 10.1108/JPIF-01-2017-0006.

Cajias, M., Freudenreich, P. and Freudenreich, A. (2020), “*Exploring the determinants of real estate liquidity from an alternative perspective: censored quantile regression in real estate research*”, *Journal of Business Economics*, Vol. 90 No. 7, pp. 1057-1086, doi: 10.1007/s11573-020-00988-w.

Čeh, M., Kilibarda, M., Lisec, A. and Bajat, B.(2018), “*Estimating the performance of random forest versus multiple regression for predicting prices of the apartments*”, *ISPRS International Journal of Geo-Information*, Vol. 7 No. 5, p. 168, doi: 10.3390/ijgi7050168.

Cetin, S., Ulgen, A., Dede, I. and Li, W. (2021), “*On fair performance comparison between Random Survival Forest and Cox regression: an example of colorectal cancer study*”, *SciMedicine Journal*, Vol. 3 No. 5, pp. 66-76, doi: 10.28991/SciMedJ-2021-0301-9.

Cirman, A., Pahor, M. and Verbic, M. (2015), “*Determinants of time on the market in a thin real estate market*”, *Engineering Economics*, Vol. 26 No. 1, pp. 4-11, doi: 10.5755/j01.ee.26.1.3905.

Cox, D.R. (1972), “*Regression models and life-tables*”, *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 34 No. 2, pp. 187-220, doi: 10.1111/j.2517-6161.1972.tb00899.x.

- Dimopoulos, T., Tyralis, H., Bakas, N.P. and Hadjimitsis, D. (2018), “*Accuracy measurement of random forests and linear regression for mass appraisal models that estimate the prices of residential apartments in Nicosia, Cyprus*”, *Advances in Geosciences*, Vol. 45, pp. 377-382, doi: 10.5194/adgeo-45-377-2018.
- Do, A.Q. and Grudnitski, G.A. (1992), “*A neural network approach to residential property appraisal*”, *The Real Estate Appraiser*, Vol. 58 No. 3, pp. 38-45.
- Fan, G.Z., Ong, S.E. and Koh, H.C. (2006), “*Determinants of house price: a decision tree approach*”, *Urban Studies*, Vol. 43 No. 12, pp. 2301-2315, doi: 10.1080/00420980600990928.
- Gabrielli, L., Ruggeri, A.G. and Scarpa, M. (2023), “*Location, location, location': fluctuations in real estate market values after COVID-19 and the war in Ukraine based on econometric and spatial analysis, random forest, and multivariate regression*”, *Land*, Vol. 12 No. 6, p. 1248, doi: 10.3390/land12061248.
- Gao, G., Bao, Z., Cao, J., Qin, A.K. and Sellis, T. (2022), “*Location-centered house price prediction: a multi-task learning approach*”, *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 13 No. 2, pp. 1-25, doi: 10.1145/3501806.
- Gröbel, S. (2019), “*Analysis of spatial variance clustering in the hedonic modeling of housing prices*”, *Journal of Property Research*, Vol. 36 No. 1, pp. 1-26, doi: 10.1080/09599916.2018.1562490.
- Harrell, F., Califf, R., Pryor, D., Lee, K. and Rosati, R. (1982), “*Evaluating the yield of medical tests*”, *Journal of the American Medical Association*, Vol. 247 No. 18, pp. 2543-2546, doi: 10.1001/jama.1982.03320430047030.
- Hoerberichts, M., van Rooij, M. and Siegmann, A. (2013), “*House list prices and durations in boom and bust*”, available at: <https://deliverypdf.ssrn.com/delivery.php?ID=843106086008086070090023068097075010017051024001008020124117114123013031071031114091049101001060018038058112064065030069080124109034011046042124019015110117120004041055094020004081108098014124083003088083001095105009068120096120006097119117122098099&EXT=pdf&INDEX=TRUE> (accessed 3 August 2023), doi: 10.2139/ssrn.2321521.
- Ishwaran, H., Kogalur, U.B., Blackstone, E.H. and Lauer, M.S. (2008), “*Random survival forests*”, *The Annals of Applied Statistics*, Vol. 2 No. 3, pp. 841-860, doi: 10.1214/08-AOAS169.
- Jaen, R.D. (2002), “*Data mining: an empirical application in real estate valuation*”, *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference, 2002, Florida*, pp. 314-317.
- Kluger, B.D. and Miller, N.G. (1990), “*Measuring residential real estate liquidity*”, *Real Estate Economics*, Vol. 18 No. 2, pp. 145-159, doi: 10.1111/1540-6229.00514.

- Kok, N., Koponen, E.L. and Martínez-Barbosa, C.A. (2017), “*Big data in real estate? From manual appraisal to automated valuation*”, The Journal of Portfolio Management, Vol. 43 No. 6, pp. 202-211, doi: 10.3905/jpm.2017.43.6.202.
- Krainer, J. (1999), “*Real estate liquidity*”, Economic Review, Vol. 3, pp. 14-26.
- Miao, F., Cai, Y.P., Zhang, Y.T. and Li, C.Y. (2015), “*Is random survival forest an alternative to Cox proportional model on predicting cardiovascular disease?*”, in Lacković, I. and Vasic, D. (Eds), 6th European Conference of the International Federation for Medical and Biological Engineering, IFMBE Proceedings, Springer, Cham, pp.740-743, doi: 10.1007/978-3-319-11128-5\_184.
- Mora-Garcia, R.T., Cespedes-Lopez, M.F., Perez-Sanchez, V.R., Marti, P. and Perez-Sanchez, J.C. (2019), “*Determinants of the price of housing in the province of Alicante (Spain): analysis using quantile regression*”, Sustainability, Vol. 11 No. 2, p. 437, doi: 10.3390/su11020437.
- Nguyen, A. (2018), Housing Price Prediction, Union College, available at: <https://cs.union.edu/Archives/SeniorProjects/2018/CS.2018/files/nguyena2/nguyena2-499-report.pdf> (accessed 3 August 2023).
- Özsoy, O. and Şahin, H. (2009), “*Housing price determinants in Istanbul, Turkey: an application of the classification and regression tree model*”, International Journal of Housing Markets and Analysis, Vol. 2 No. 2, pp. 167-178, doi: 10.1108/17538270910963090.
- Pérez-Rave, J.I., Correa-Morales, J.C. and González-Echavarría, F. (2019), “*A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes*”, Journal of Property Research, Vol. 36 No. 1, pp. 59-96, doi: 10.1080/09599916.2019.1587489.
- Shahhosseini, M., Hu, G. and Pham, H. (2020), “*Optimizing ensemble weights for machine learning models: a case study for housing price prediction*”, Smart Service Systems, Operations Management, and Analytics: Proceedings of the 2019 INFORMS International Conference on Service Science, 2020, Springer International Publishing, pp. 87-97.
- Smith, B.C. (2010), “*Spatial heterogeneity in listing duration: the influence of relative location to marketability*”, Journal of Housing Research, Vol. 18 No. 2, pp. 151-171, doi: 10.1080/10835547.2009.12092005.
- Statistisches Amt der Landeshauptstadt München (2022), “*Demografie der Münchner Bevölkerung 2021*”, Münchner Statistik, Vol. 1.
- Tay, D.P. and Ho, D.K. (1992), “*Artificial intelligence and the mass appraisal of residential apartments*”, Journal of Property Valuation and Investment, Vol. 10 No. 2, pp. 525-540, doi: 10.1108/14635789210031181.

Yavas, A. and Yang, S. (1995), “*The strategic role of listing price in marketing real estate: theory and evidence*”, Real Estate Economics, Vol. 23 No. 3, pp. 347-368, doi: 10.1111/1540-6229.00668.

Yoo, S., Im, J. and Wagner, J.E. (2012), “*Variable selection for hedonic model using machine learning approaches: a case study in Onondaga County, NY*”, Landscape and Urban Planning, Vol. 107 No. 3, pp. 293-306, doi: 10.1016/j.landurbplan.2012.06.009.

## Appendix

**Table 4.3** Non-geographic features - dwelling characteristics

<b>Variable name</b>	<b>Unit</b>
1. Price	€
2. Living area	m <sup>2</sup>
3. Rooms	number
4. Floor	number
5. Bathtub	binary
6. Balcony	binary
7. Terrace	binary
8. Elevator	binary
9. Construction year	year

## **5 Conclusion**

This section provides a summary of the three papers in this dissertation including the research objective, the study design, as well as the key findings and its implications. The dissertation concludes with some final remarks on the core topics and an outlook on further research in the field of liquidity of residential real estate.

### **Executive summaries**

**Paper 1:        Liquidity and prices: A cluster analysis of the German residential real estate market**

---

#### **Problems and objective**

The first paper of this dissertation sheds light into the under-researched German residential real estate market by analysing prices and liquidity on the investment and rental market separately for 380 regions in Germany. On the residential real estate market, the process of selling or renting out an apartment comprises of two essential components. The first component is the introduction of the apartment onto the market at a price determined by the seller or landlord. The second component is the time it takes until a prospective buyer or tenant is willing to take the dwelling off the market and to pay the price. Contingent upon a matching of the expectations of supply and demand, a market can operate. The easier, thus faster this matching occurs, the higher the liquidity on the market. Currently, the assessment of the real estate market, especially in Germany, is mainly based on the consideration of prices, see e.g. Bauer et al. (2013) and an de Meulen and Mitze (2014), among others. Hence, the price development is captured by a variety of indices worldwide. However, there exist no liquidity indices capturing the time on market development on the residential real estate market. Nevertheless, considering the of time on market development seems to be crucial for detecting tight market conditions. To improve the assessment of the residential market, this article provides quality- and spatial-adjusted liquidity indices for the residential investment and rental market, as complementary demand indices, separately for 380 regions in Germany. Hence, the objective of this article is to provide deeper insights to the German residential real estate market by combining the consideration of price and liquidity indices.

#### **Methodology and data**

Price and liquidity indices are developed for the investment and rental market separately for 380 of 401 NUTS 3 regions in Germany from 2013 Q1 to 2018 Q4. The time dummy approach is applied to generate price and liquidity indices for the German residential investment and rental market. The coefficients of the time dummy variable are extracted from a quality- and spatial-

adjusted hedonic regression model. While the General Additive Model for Location, Scale and Shape (GAMLSS) introduced by Rigby and Stasinopoulos (2005) is used to generate the price indices, the Cox (1972) Proportional Hazards Model is applied to generate the liquidity indices. Applying the “Partitioning Around Medoids (PAM)” clustering algorithm, the regions are clustered with respect to their price and liquidity development after the optimal number of clusters was found by means of the average silhouette method. Economic and socioeconomic data is applied to further analyse the clusters and to identify similarities. Furthermore, the clusters are interpreted from a geographic perspective. Therefore, microdata is gathered for more than 4.5 million observations, split into about 1.5 million on the investment market and about 3 million on the rental market.

### **Results and their contribution to science and practice**

Generally, the price and liquidity development are geographically quite differently distributed across Germany. Flourishing regions are mainly the major cities and their surrounding conurbations as well as economically strong regions in southern and north-western parts of Germany. The opposite is to observe in rural regions in the eastern parts as well as structurally weak regions in the western parts. The clusters are analysed by means of further economic and socioeconomic data to identify similarities. Furthermore, the clusters are interpreted from a geographic perspective. The results show that the allocation to cluster 1 is always supported by higher growth rates in the variables, population, working population and real GDP, implying higher demand for space. Moreover, in each of the four categories, cluster 1 reveals a lower unemployment rate as well as a higher disposable income. One of the most interesting implications is, that apparently a large part of the German population has developed into professional real estate investors. In Germany the largest share of landlords is the one of the so-called non-professional landlords. As the regions assigned to cluster 1, displaying the most significant price increase, seem to be chosen based on a very sophisticated market analysis by identifying the regions with the strongest fundamental data, it seems like the dominating market players have significantly increased their knowledge and approach for investing in residential real estate. To summarize, the findings of this article contribute to a deeper understanding of the German residential real estate market by the introduction of a liquidity index. Furthermore, the importance of a combined analysis of price and liquidity is emphasized. The results can be used to deduce investment strategies and to assist policy makers on the identification of tight markets and a prioritization of subsequent actions.

---

**Paper 2: Exploring the determinants of real estate liquidity from an alternative perspective: censored quantile regression in real estate research**

---

**Problems and objective**

The aim of the second paper of this dissertation is a detailed understanding of the major determinants of time on market of rental dwellings along the entire liquidity distribution. The most established approach analyzing time on market is the Cox (1972) proportional hazards model (PHM). This model implies strong restrictions on the variation of estimated coefficients, i.e. the proportional hazards assumption needs to be fulfilled. However, we assume the direction and magnitude of the effect of covariates on real estate liquidity to vary along the liquidity distribution. Therefore, this study introduces a modelling technique which is new to the field of real estate liquidity analysis, the Censored Quantile Regression (CQR) model. The decisive feature is that CQRs are able to model any quantile of the distribution of the dependent variable time on market. According to Chaudhuri et al. (1997) this is as a great advantage compared to mean regressions, as distributions might not only be different in terms of their means but might differ especially in their upper and lower tails. CQRs can quantify the impact of a covariate on the dependent variable for any quantile, compared to only the center of the population. The results are used to validate the outcome of the Cox (PHM). The CQR, as an expansion of the survival regression analysis, is expected to yield more accurate estimations and to provide a more complete statistical analysis for understanding the factors driving liquidity and the underlying demand.

To the best of the authors' knowledge this is the first paper to apply a censored quantile regression approach in real estate research to investigate the determinants of time on market.

**Methodology and data**

The CQR model aims to explain the variation of time on market as a function of dwelling characteristics and other spatial and socioeconomic characteristics. The Quantile Regression (QR) approach goes back to Koenker and Bassett Jr. (1978). It allows the estimation of the impact a covariate has on the dependent variable, conditional on each specific quantile of the dependent variable along the distribution. Zietz et al. (2008), Farmer and Lipscomb (2010), Mak et al. (2010), Liao and Wang (2012), among others as well as an de Meulen and Mitze (2014) and Tomschke (2015) for the German market applied the QR approach in real estate pricing. An important feature of the time on market as a dependent variable is that some observations do not change their event status throughout the observation period. This means, that some dwellings remain on the market after the end of the observation period. In this case the response variable, time on market, is right-censored. To deal with censoring within the QR framework, Powell's (1984, 1986) approach, addressing fixed censoring, is applied to this study.

The dataset consists of 482,196 observations on the rental markets of the seven largest German cities (descending order by population: Berlin, Hamburg, Munich, Cologne, Frankfurt, Stuttgart, Dusseldorf) from 2013 Q1 to 2017 Q4. As the impact and magnitude of a coefficient might be different in different locations across the country each city is analyzed individually.

### **Results and their contribution to science and practice**

The results show that the magnitude and direction of impact of an explanatory variable on time on market differs between the cities. Consequently, it is important to analyse each city individually and not the seven cities or even the German market as whole. For most covariates, consistent signs of the regression coefficients were found across the quantiles along the time on market distribution. However, for some covariates the magnitude and direction of effect on the liquidity of a dwelling exhibits differences between time on market quantiles within a city. This implies that, the proportional hazards assumption, underlying the Cox PHM is violated for specific explanatory variables and cities and hence emphasizes the application of the CQR approach for the time on market analysis. The study concludes that the heterogeneity across the liquidity quantiles, as well as the heterogeneity between the cities, are accountable for the distinguishable impacts of changes in the covariates on time on market. These findings emphasize the importance of market segmentation for a more detailed analysis and understanding of the rental real estate market. This article extends the existing literature on residential liquidity analysis by the initial application of the advanced Censored Quantile Regression (CQR). These findings should be of interest to current and future landlords, as they reveal both the characteristics of dwellings along the liquidity distribution (e.g. the existence of built-in kitchen), as well as the impact of a change in characteristics on the liquidity of dwellings (e.g. impact of installing a built-in kitchen). Therefore, landlords should be able to infer whether a dwelling displays the characteristics of a highly liquid thus highly demanded dwelling, or which actions they could take to increase the marketability, e.g. install a built-in kitchen or change the floor plan to increase the number of rooms. Furthermore, the findings suggest that nationwide or even statewide policy measures might not be suited to address the specific circumstances on the residential rental real estate market of a particular region, city or neighbourhood.

### **Paper 3:       What are tenants demanding the most? A machine learning approach for the prediction of time on market**

---

#### **Problems and objective**

The availability of large amounts of data on the real estate market combined with sophisticated statistical techniques yields exciting possibilities for highly accurate estimations. One emerging

approach to deal with big data is machine learning. Since the end of the twentieth century the application of machine learning techniques to the field of real estate research has gained great interest. Several articles have been published applying different machine learning techniques such as neural networks (see e.g. Jean (2002)), decision trees (see e.g. Fan et al. (2006)), or ensemble methods (see e.g. Kok et al. (2017), Čeh et al. (2018)). For the prediction of real estate prices, machine learning techniques have developed as an upcoming alternative to the common hedonic models based on multiple regression. However, to the best of the authors' knowledge, there exists no application of machine learning techniques for the analysis of the time on market in the context of real estate market analysis. Hence, the aim of this paper is to contribute to the current state of research on the real estate market by applying a machine learning algorithm, specifically the random survival forest algorithm, to the analysis of time on market on the residential rental market in the city of Munich. The most important predictors of time on market are revealed and it is analyzed how the survival probability of residential rental apartments responds to these major characteristics.

### **Methodology and data**

In this paper the time on market on the residential rental market of Munich is analyzed. As the variable of interest is the time span an apartment is up for rent on the market, a duration is estimated. Therefore, another modelling class named survival models is needed. The most popular approach to analyze the time on market is the Cox (1972) proportional hazards model (PHM). In this paper the random survival forest approach, based on decision trees, is introduced to the real estate market. To the best of the authors' knowledge this is the first paper, to apply the random survival forest approach to the liquidity analysis on the real estate market. The most important predictors of time on market are revealed and it is analyzed how the survival probability of residential rental apartments responds to these major characteristics. The underlying dataset comprises a sample of 115,264 apartments based on micro data on the residential rental market of Munich over a period of nine years from 2013 Q1 to 2021 Q4. To precisely estimate the marketing time several covariates precisely describing the apartment as well as the surrounding environment are included.

### **Results and their contribution to science and practice**

Results show that by far the strongest influence on the marketing time of an apartment has the price of the apartment, followed by the living area, the construction year and the year of listing. The next category of influential variables are distance variables. The distance of a dwelling to the next hairdresser, the next bakery, the city center, the nearest school, and the next windmill impact the time on market. Furthermore, the seasonal aspect of renting out an apartment affected the time

on market. The estimation results were interpreted using partial dependence (PD) plots that visualize the relationship the machine learning algorithm detects between the dependent variable time on market and its determining factors in the underlying dataset. The time on market for an apartment in Munich is lowest at a price of 750 € per month, an area of 60 m<sup>2</sup>, built in 1985; is in a range of 200–400 meters from the important amenities and is located about 4.5 kilometers from the city center. Moreover, the time on market is lowest for apartments listed on the market in April or June. Generally, in an annual comparison, apartments listed in 2019 were the fastest to be rented out.

The insights are interesting for further research as well as for practical implications. Real estate managers can derive their strategies regarding their portfolio management and can adjust their portfolio according to the predicted survival probabilities. With a profound understanding of the time on market, its determinants, and its behavior to changes in the determinants, real estate professionals as well as private investors can minimize their cash-flow failure. With this detailed time on market assessment, current and future landlords of apartments in Munich should be able to infer how long it will take to let their apartments or which actions to take to advance the letting process.

### **Final remarks**

During the last years the German residential real estate market has been characterized by increasing demand for residential space combined with low mortgage rates leading to rising prices as well as rents and way too little housing supply especially in the metropolises. With the improvement in computational power and the ability to process large amounts of data new fields of research have become accessible especially when it comes to the application of advanced econometric methods. Several articles have been contributing to a more sophisticated analysis of the German real estate market, especially for the estimation and prediction of housing prices. The objective of this dissertation is to contribute to the analysis of the German residential real estate market, especially to deliver insights about the second integral component in the marketing process of residential real estate, namely the liquidity. Liquidity of residential real estate is defined as the inverse of the time an apartment is on the market. The aim of the first article is to provide a combined analysis of price and liquidity indices separately for 380 regions on the German investment and rental market. While price indices are a common tool for the real estate market assessment, this is the first paper, to the best of the authors' knowledge, to introduce a liquidity index to the field of residential real estate research. The second article of this dissertation builds upon these findings and hence the aim of this paper is to derive a detailed understanding of the major drivers of marketing time of rental dwellings individually for the seven largest German cities. The article introduces a modelling technique, namely Censored Quantile Regressions (CQR), to investigate the impact of any covariate on the time on market along the entire time on

market distribution, instead of just the center of the population. In the third article of this dissertation a machine learning approach is introduced to the liquidity analysis on the real estate market. The article empirically analyzes the most influential variables that affect the time on market of rental apartments for the city of Munich. To the best of the authors' knowledge this is the first paper, to apply the random survival forest approach to the liquidity analysis on the real estate market. In summary, this dissertation contributes to a deeper and profound understanding of the German residential real estate market, especially of the residential liquidity. All three articles approach the residential liquidity topic with different statistical methods and from different point of views that provide new insights from scientific as well as practical perspectives. The insights about residential liquidity delivered in this dissertation might be an illuminating addition to the existing research about real estate price predictions for a deeper understanding of the residential real estate market. The findings are interesting for further research as well as for practical implications. A profound understanding of the time on market on the residential real estate market in Germany is of great importance for private, institutional, and governmental market players. It is essential to evaluate market fundamentals, to assess and forecast cyclical movements as well as to derive policy implications.

Since the aim of this dissertation is to shed more light into the liquidity analysis on the German residential real estate market, there are several questions and issues which might be addressed with further research. The application of modern machine learning algorithms promises interesting results and deeper insights to the liquidity analysis. Therefore, the third article of this dissertation could be a fruitful starting point. As this article only refers to the city of Munich, an extension to and comparison of the findings to other markets could deliver further insights to the time on market behaviour. It might be interesting whether the most influential variables and their impact on the time on market vary between different markets. Furthermore, commonalities and differences of urban and rural residential demand could be investigated via the respective marketing time. Following Gabrielli et al. (2023) a comparison at different time points could yield further insights into the reaction of time on market to changes in its dwelling specific and neighbourhood specific factors. Thereby it could be evaluated which impact economic and socioeconomic events have on the time on market. With the increase in mortgage rates, the changes in economic situation and non the less the availability of AI and sophisticated machine learning algorithms the residential real estate market remains an exciting and inspiring field of research.

---

**References**

- an de Meulen, P. and Mitze, T. (2014). Exploring the spatial variation in quality-adjusted rental prices and identifying hot spots in Berlin's residential property market. *Regional Studies, Regional Science*, 1(1), 310-328.
- Bauer, T.K., Feuerschuetz, S., Kiefer, M., an de Meulen, P., Micheli, M., Schmidt, T. and Wilke, L. H. (2013). Ein hedonischer Immobilienpreisindex auf Basis von Internetdaten: 2007–2011, *ASTA Wirtschafts-und Sozialstatistisches Archiv*, 7(1-2), 5-30.
- Čeh, M., Kilibarda, M., Lisec, A. and Bajat, B. (2018). Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *ISPRS International Journal of Geo-Information*, 7(5), 168.
- Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics*, 25(2), 715-744.
- Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–202.
- Fan, G.Z., Ong, S.E. and Koh, H.C. (2006). Determinants of house price: a decision tree approach. *Urban Studies*, 43(12), 2301-2315.
- Farmer, M. and Lipscomb, C. (2010). Using quantile regression in hedonic analysis to reveal submarket competition. *Journal of Real Estate Research*, 32(4), 435-460.
- Gabrielli, L., Ruggeri, A.G. and Scarpa, M. (2023). 'Location, Location, Location': Fluctuations in Real Estate Market Values after COVID-19 and the War in Ukraine Based on Econometric and Spatial Analysis, Random Forest, and Multivariate Regression. *Land*, 12(6), 1248.
- Jaen, R.D. (2002). Data mining: an empirical application in real estate valuation. Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference, 2002, Florida, 314-317.
- Koenker, R. and Bassett, G.W. Jr (1978). Regression quantiles. *Econometrica, Journal of the Econometric Society*, 46, 33–50.
- Kok, N., Koponen, E.L. and Martínez-Barbosa, C.A. (2017). Big data in real estate? From manual appraisal to automated valuation. *The Journal of Portfolio Management*, 43(6), 202-211.
- Liao, W.C. and Wang, X. (2012). Hedonic house prices and spatial quantile regression. *Journal of Housing Economics*, 21, 16-27.

Mak, S., Choy, L. and Ho, W. (2010). Quantile regression estimates of Hong Kong real estate prices. *Urban Studies*, 47(11), 2461-2472.

Powell, J.L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25(3), 303-325.

Powell, J.L. (1986). Censored regression quantiles. *Journal of Econometrics*, 32(1), 143-155.

Rigby, R.A. and Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507-554.

Thomschke, L. (2015). Changes in the distribution of rental prices in Berlin. *Regional Science and Urban Economics*, 51, 88-100.

Zietz, J., Zietz, E.N. and Sirmans, G.S. (2008). Determinants of house prices: a quantile regression approach. *The Journal of Real Estate Finance and Economics*, 37, 317-333.