Censored Quantile Regressions and the Determinants of Real Estate Liquidity

Marcelo Cajias[a,*], Philipp Freudenreich[b], Anna Heller[c], Wolfgang Schäfers[d]

[a] *Patrizia Immobilien AG, Fuggerstr. 26, 86150 Augsburg, Germany, Phone +49 176 80293 102.*
[b] *University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany, Phone +49 941943 5091.*
[c] *University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany, Phone +49 941943 5688.*
[d] *University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany, Phone +49 941 943-5071.*

Abstract

In this paper we analyze the liquidity (time on market) of rental dwellings and its determinants for different liquidity quantiles in the largest seven German cities. The determinants are estimated using censored quantile regressions in order to investigate the impact on very liquid to very illiquid dwellings. We use micro data to examine the time on market for about 400,000 observations from the first quarter of 2013 to the second quarter of 2017. Hedonic and socioeconomic characteristics as well as spatial gravity variables and various fixed effects are used as exogenous determinants. For almost all regression coefficients we find consistent signs across all quantiles of the time on market distribution, i.e. the proportional hazard assumption, underlying the Cox model, is not violated. However, we find substantial differences in the magnitude of the regression coefficients over the liquidity quantiles. This is the first paper, to the best of our knowledge, to apply censored quantile regressions to the liquidity analysis on the real estate market.

Keywords: Real estate liquidity, Duration analysis, Time on market, Censored quantile regressions, Cox-hazard, German housing.

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

*Corresponding author.
Email addresses: `marcelocajias@hotmail.com` (Marcelo Cajias), `philipp.freudenreich@ur.de` (Philipp Freudenreich), `anna.heller@ur.de` (Anna Heller), `wolfgang.schaefers@ur.de` (Wolfgang Schäfers)

1. Introduction

Heterogeneity is an omnipresent conception in the world of real estate. Even if two properties are identically constructed, their non-duplicable location makes them inimitable. This uniqueness impedes the matching between seller and buyer and often affects the pricing as well as the time the transaction requires. Both pricing of real estate sales as well as the associated time on market have generated a notable amount of literature.

Heterogeneity can not only be found for the characteristics of the property but for markets respectively market segments. As depicted in figure 1, the time on market for German rental real estate experienced a continuous contraction within the last years. The non-censored subsample of lease transactions reveals an irregular but altogether constant decline in the median time on market from the second quarter of 2014 onwards. Besides the fourth quarter of 2014 a narrowing distribution of time on market is observed as well. For this study the inverse of the time on market of rental dwellings is used to construct a liquidity measure based on the liquidity definition by Wood and Wood (1985). By splitting the sample into seven quantiles based on their liquidity level, this study is the first to the best of our knowledge to investigate the determinants of time on market for individual market segments by applying a censored quantile regression for an extensive dataset covering the largest seven German cities.

The study is able to identify patterns for the impact the explanatory variables have on time on market. Those patterns vary based on the location of the dwelling as well as the liquidity level of the dwelling, which is defined by allocating the dwelling to a certain time on market quantile. While the direction of the impact a change in an explanatory variable has on the liquidity of a dwelling changes for the various variables and cities, it is quite distinct that the magnitude of those effects rises with the level of illiquidity for all variables and cities. The study concludes that the heterogeneity within the liquidity of dwellings is accountable for significantly different interpretations of the effect a change in the explanatory variables has on the time on market of dwellings.

If this heterogeneity is not taken into account liquidity estimates are likely to be biased as well as market knowledge might be distorted. Deeper insights into specific market segments do not only provide reduced search costs to households (see e.g. Malpezzi (2003), Goodman and Thibodeau (2007)) but also improve the financial appraisal of creditors and investors, private or institutional. Therefore the segmentation into submarkets leads to a more profound understanding of liquidity patterns on the residential real estate market. One of the first empirical studies analyzing real estate liquidity for different market segments was conducted by Belkin et al. (1976). They define submarkets

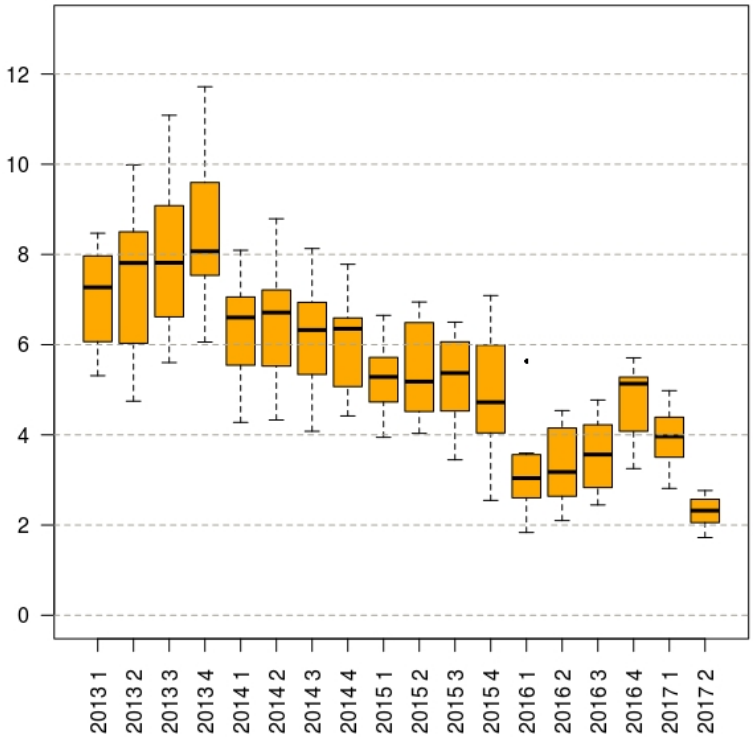

Figure 1: Distribution of time on market in weeks in Germany

according to geographic areas, price segments and buyers' search space and analyze the relationship between TOM and the spread between listing price and selling price using OLS estimation techniques. They find essential differences by market segments. Especially in high-price submarkets deviations from the initial list price have a more pronounced effect on ToM. The determinants of ToM considering different price segments are further analyzed by Kang and Gardner (1989), Kalra and Chan (1994), Yavas and Yang (1995) and Allen et al. (2009) among others. Kang and Gardner (1989) find that the impacts on ToM do vary in magnitude between the low-, medium- and high-price segments. While Kang and Gardner (1989), applying an OLS estimation, did not identify the simultaneity problem between ToM and the selling price, Yavas and Yang (1995) apply a 2SLS estimation to capture the fact that ToM and price mutually influence each other. They exhibit a significant positive impact of price on ToM in the medium-price subgroups whereas this effect is insignificant for houses in the low- and high-price segments. Allen et al. (2009) analyze the relationship between asking rent and ToM on the rental market of single family residential rental listings using a multi-step procedure. ToM is estimated via a hazard model with a

Weibull distribution. According to asking rents the sample is divided into three price subgroups. They find that underpricing of asking rents and TOM move in the same direction in every price segment, however the effect being stronger in the medium- and high-price submarket compared to low-price houses. As a conclusion from these findings, market segmentation seems to be a valuable contribution for understanding liquidity patterns. This is also confirmed by Guasch and Marshall (1985), Ong and Koh (2000), Turnbull et al. (2006) and McGreal et al. (2009) among others, who segment markets according to the number of rooms, the number of units in a structure, the geographical region, the property type or by market cycle. More closely related to our paper is the article of Turnbull and Dombrow (2006), as they divide their sample into low-, medium- and high-liquidity submarkets. They explore the impact of listing density on time on market for the pooled sample, for different market cycles and for different market cycles combined with different liquidity segments. Applying a 3SLS estimation, they find that the significance as well as the magnitude and directions of the impact of the spatial competition variables on time on market vary between the different liquidity submarkets. Following these findings, we hypothesize that

1. different factors play a role in determining liquidity in low-, medium- and high-liquidity segments respectively and
2. the direction and magnitude of these factors vary across liquidity segments.

As we are able to investigate and compare the impact factors on the time on market for the seven largest German cities, we additionally hypothesize that

3. the direction and magnitude of these factors vary across these cities.

Nowadays, the most popular model for the estimation of duration data is the Cox (1972) proportional hazards models (PHM) or also commonly used is the accelerated failure time model. However, in terms of the econometric model, our paper strongly differs from the preceding studies as we apply censored quantile regressions (CQRs) to real estate liquidity analysis. While median regression, which is a special case of QR, goes back as far as 1760 (see Stigler (1984)), QR has been introduced by Koenker and Bassett Jr. (1978). Compared to the accelerated failure time model or the Cox (1972) 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 for our 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 by their means but might especially

differ in their upper and lower parts. 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. Yu et al. (2003) exhibit that QR models are strongly associated to the three statistical concepts of regression, robustness and extreme value theory.

Since its introduction the QR approach has received increasing attention, theoretically as well as empirically, and has been applied to many different research areas.[1] In the real estate literature, more precisely in the area of hedonic pricing, QRs have been applied by Zietz et al. (2008), Farmer and Lipscomb (2010), Mak et al. (2010), Liao and Wang (2012) among others. However, when it comes to real estate liquidity, we are the first, to the best of our 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) have initially, as well as Lüdemann et al. (2006), Fitzenberger and Wilke (2010), Schmillen and Möller (2010) among others, have lately applied CQRs. Conceptually our analysis is highly related to Lüdemann et al. (2006). To analyze censored survival data there are three common approaches introduced by Powell (1984, 1986), Portnoy (2003) as well as Peng and Huang (2008). A comprising description on the implementation of the three aforementioned approaches into the R package 'quantreg' can be found in Koenker (2008). Portnoy (2003) and Peng and Huang's (2008) approach for random censoring can be applied, if censoring values are only known for the censored observations. In contrast, Powell's (1984,1986) approach for fixed censoring is used, if the censoring values are observed for all observations and thus, is best suited for our analysis. In particular, the CQR method used in this paper goes back to Koenker and Bilias (2002).
The remainder of this paper proceeds as follows: The next section describes the underlying econometric model, followed by a detailed description of the dataset and the descriptive statistics in section 3. Estimation results are presented and discussed in section 4. Section 5 concludes.

[1] for survival analysis see e.g. Crowley and Hu (1977), Yang (1999), Koenker and Geling (2001)
for medical research see e.g. Cole and Green (1992), Royston and Altman (1994), Harder et al. (2005), Owen et al. (2005), Wei et al. (2006), Beyerlein et al. (2008), Wehby et al. (2009)
for financial economics see e.g. Taylor (1999), Bassett Jr. and Chen (2001)
for environmental research see e.g. Hendricks and Koenker (1992), Pandey and Nguyen (1999)
for labour economics see e.g. Rose (1992), Buchinsky (1994, 1995)
for a review on the application fields of QRs see Yu et al. (2003)

2. Econometric model

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 our 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 in the market into being out of the market in weeks, i.e. time on market (ToM). 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 occured until a certain time t and is fomally defined as

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

with $f(x)$ being the probability density function (p.d.f.) of the time untill 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

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

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 simple 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 most importantly focusses on the estimation of the hazard function by transforming the response variable in units of time into a discrete variable, so called conditional odds. The Cox regression for a specific observation i is given as

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

where x is a vector of covariates (without the constant), β 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 $\lambda_0(t)$. It assumes, however, proportional hazards, meaning that the hazard function is a constant

function of time. Taking logs, we get a simple additive model for the log of the hazard rate

$$\log \lambda(t|x) = \log \lambda_0(t) - x'\beta, \tag{4}$$

and thus, the conditional survival function $S(t|x)$ can be described as

$$\log(-\log S(t|x)) = \log \Lambda_0(t) - x'\beta. \tag{5}$$

Consequently, the model can be written as

$$\log \Lambda_0(T) = x'\beta + u, \tag{6}$$

what is equivalent to the transformation model

$$h(T) = x'\beta + u, \tag{7}$$

with $h(T)$ being a monotone transformation of the observed survival time T and u being iid with extreme value distribution $F(u) = 1-\exp(-\exp(u))$. Hence, the Cox (1972) PHM can be written as a monotone transformation of the observed survival time T linearly depending on the covariates x plus iid error u. The elapse of time that a dwelling is offered in the market corresponds to an event that can be censored on the left or the right. Censoring refers to incomplete event cases in which the begining or the end of an event is unknown at the end of the observation period. While left censoring is rarely observed in studies, right censoring is more common and arises in our study when the landlord doesn't change the status of the dwelling in the Multiple Listing Services (MLS) database or the dwelling is still being offered in the market. In order to overcome censoring, the decomposition of the events in units of time or conditional odds is one of the main benefits of the Cox regression framework since it allows the censored events of the sample to contribute to the model until the end of the observation period.

While the Cox (1972) 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 over the last decade to account for conditional survival functions across different levels of the response. The traditional Cox regression estimates the conditional survival function for the entire sample based on the assumption of homoscedasticity within the sample. In other words, the covariates are expected to exert the same impact on the response regardless of the distribution of the response, e.g. high liquid and low liquid dwellings. While this approach is absolutely correct, it ignores conditional elasticities, meaning

that for example highly liquid dwellings respond differently to certain covariates than very illiquid dwellings. In this context, the quantile regressions have arisen as a method for estimating conditional regressions within the sample as a function of the quantile distribution of the response. In this paper, we aim at employing a unique thechnique corresponding to the survival quantile regression which has not been employed in the context of real estate liquidity, to the best of our knowledge. This method, introduced by Koenker and Bassett Jr. (1978), yields a robust and more flexible alternative for the estimation of parametric and semiparametric duration models, imposing less distributional assumptions. Moreover, there are no imposed modelling assumptions to be empirically proven true and thus, misspecification of the model is less likely. In contrast, the Cox (1972) PHM must e.g. satisfy the proportionality assumption, implying constant hazard ratios over time. Another advantage over the Cox (1972) PHM is the straightforward interpretation of regression coefficients on the dependent variable, wich provides a deeper specification of the data. Therefore, the use of QRs is highly appropriate for this analysis.

The origin of the QR model goes back to Koenker and Bassett Jr. (1978). It is a location model estimating the relationship of the covariates x with the dependent variable y, conditional on the quantile τ of y. The quantile $\tau \in (0,1)$ is defined as the value of y that separates the observations into the fraction τ below and the fraction $1-\tau$ above. Thus, the quantile τ of a random variable Y is defined as the minimum value q_τ, so that

$$q_\tau = F^{-1}(\tau) = inf(y : F(y) \geq \tau), \tag{8}$$

where $F(y) = P(Y \leq y)$ denotes the cumulative distribution function (c.d.f.) of Y. Hence, the median for example is described by $\tau = 0.5$.

Following Doksum and Gasko (1990), a lot of survival analysis models, such as the PHM, the proportional odds model or the accelerated failure time model, can be linked to the general transformation model

$$h(y_i) = x_i'\beta + u_i. \tag{9}$$

$h(y_i)$ denotes a monotone transformation of the observed dependent variable y_i, linearly depending on a $k \times 1$ vector of covariates x_i with $x_{1i} \equiv 1$ and an iid error u_i. With the error term u being defined as $u^\tau \equiv h(y) - x'\beta^\tau$ it follows that $Quant_\tau(u^\tau|x) = 0$. Thus, the conditional quantile function of the transformed

dependent variable y_i can be denoted as

$$Quant_\tau(h(y_i)|x_i) = x_i'\beta^\tau. \tag{10}$$

It describes the family of QR models. β^τ denotes a $k \times 1$ vector of regression parameters dependent on the quantile τ.

Applying the 'log'-transformation of T_i, $h(T_i) = \ln T_i$ (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

$$\ln T_i = x_i'\beta^\tau + u_i^\tau. \tag{11}$$

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

$$Quant_\tau(\ln T_i|x_i) = x_i'\beta^\tau, \tag{12}$$

where $Quant_\tau(\ln T_i|x_i)$ represents the τth conditional quantile of $\ln T_i$ given x_i. Using QRs we can investigate changes in the relation between the covariates and the time on market depending on the liquidity segment. In other words, we can estimate the factors that contribute to or restrict the survival of rental dwellings in the largest seven German cities across different levels of liquidity, e.g. highly liquid dwellings or less liquid dwellings. The quantile approach seems, furthermore, plausible given the large datasets ranging from 14,940 observations in Stuttgart to 140,244 observations in Berlin and especially due to the spatial heterogeneity in the data. The QR approach is expected to provide us with deeper insights on the underlying determinants of time on the market in the German rental housing market of the top seven cities.

2.1. Properties of the quantile regression model

Koenker and Geling (2001) note that, given

$$P(Y < y|x) = P(h(Y) < h(y)|x), \tag{13}$$

it follows, that for any monotone function $h(.)$, the QR model is invariant to monotone transformations, so that

$$Quant_\tau(h(y)|x) = h(Quant_\tau(y|x)). \tag{14}$$

Due to this essential feature, the conditional quantile functions for the untransformed duration T_i of the underlying accelerated failure time model can be written as

$$Quant_\tau(T_i|x_i) = \exp(x_i'\beta^\tau). \tag{15}$$

Another important property of the QR model, with regard to its interpretation, is the formulation of treatment effects. Dating back to Lehmann and D'abrera (1975) and Doksum (1974), the two-sample treatment response model provides the simplest formulation. Koenker and Bilias (2002) also consider the treatment effects by means of the simplest bivariate case,

$$Quant_\tau(\ln T_i|x_i) = \beta_0^\tau + \beta_1^\tau x_i, \tag{16}$$

where $x_i = 1$, if observation i received treatment and $x_i = 0$ for the control group. For a continuous treatment variable the effect β_1^τ of increasing x_i by a marginal unit from x_{i0} to $x_{i0} + \Delta$ is equal to the effect of changing x_{i1} to $x_{i1} + \Delta$. The crucial point to notice is that contrary to the traditional regression models the treatment effect is only the same within a quantile τ but might differ across quantiles.

2.2. Censored quantile regression model

An important feature of survival analysis is, that some observations do not change their event status throughout the observation period. For our analysis this means that some dwellings remain available in the MLS database by 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 our dataset Powell's (1984, 1986) approach is best suited as it adresses fixed censoring. In contrast to the Cox (1972) PHM, for QRs with fixed censoring it is necessary to know the observation specific censoring value C_i for all observations. If an observation i is censored, we cannot observe the actual survival time T_i^*, but do 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 $\hat{\beta}^\tau$ is the value of β^τ solving the minimization problem of the distance function

$$Q_N(\beta;\tau) \equiv \frac{1}{N}\sum_{i=1}^{N} \rho_\tau(\ln T_i - \min(x_i'\beta^\tau, C_i)). \tag{17}$$

The minimization term becomes $x_i'\beta^\tau$ if $x_i'\beta^\tau < C_i$ and is C_i otherwise. Thus, $x_i'\beta^\tau$ is censored from above at the upper threshold C_i. The 'check-function' $\rho_\tau(u)$ is defined as

$$\rho_\tau(u) = \begin{cases} \tau \cdot |u| & u \geq 0 \\ (1-\tau) \cdot |u| & u < 0 \end{cases} . \tag{18}$$

$\tau \cdot |u|$ denotes the penalty for underprediction and $(1-\tau) \cdot |u|$ for overprediction. The estimator $\hat{\beta}$ 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 $\hat{\beta}^{0.5}$. The coefficients can be interpreted as the change in the dependent variable that, ceteris paribus, arises from a marginal incraese in the respective regressor while keeping the dependent variable in the same quantile (see Machado and Mata (2000)). In more practical words: An increase of an explanatory variable e.g. "price" by a marginal unit, ceteris paribus, prolongs or shortens the time on market by $|1 - \exp(\hat{\beta}^\tau_{price})| * 100\%$, while the time on market remains in the same quantile τ. A prolongation of the time to event occurs if the hazard ratio $\exp(\hat{\beta}^\tau_{price})$ is greater than 1 and a reduction of the time to event if the hazard ratio is smaller than 1.
Powell (1986) demonstrates that under appropriate conditions for a certain value of τ the censored regression quantile estimator β^τ is $\sqrt{N}$-consistent and asymptotic normality is proven true if the appropriate assumptions hold for each $\tau \in \{\tau_1, ..., \tau_J\}$. While in the uncensored QR problem the objective function to be minimized $Q_N(\beta; \tau)$ is convex this nice property is not given for the censored case, leading to some strong computational difficulties.

3. Data and descriptive statistics

The estimation sample is composed of three merged data sets containing information of 394,366 observations on the rental market in the largest German cities from the first quarter of 2013 to the second quarter of 2017. Additionally, further variables are calculated with ArcGIS. Information on the rental dwellings are gathered from various Multiple Listing Services (MLS) as collected from the Empirica Systems Database, containing real estate market data from more than 100 sources, among them the most important MLS providers. Characteristics of the rental dwellings contain the time on market as the number of weeks the flat was listed in the MLS calculated by the start and end date, the asking rent in euro per sqm per month, some other housing attributes and binary variables

like "with balcony", being 1 if the flat exhibits a balcony and 0 otherwise. The degree of overpricing (DOP) as introduced by Anglin et al. (2003) is defined as the percentage deviation of the list price of an individual property from its empirically estimated market list price. Since the data is georeferenced, two spatial gravity indicators measuring the Euclidian distance of each dwelling to the geographical centroid of the ZIP and NUTS3 polygon in kilometres are incorporated. NUTS3 regions correspond to the "Nomenclature of territorial units for statistics", which is a hierarchical system for dividing up the economic territory in Europe. The NUTS3 regions cover small regions similar to counties or administrative districts. In our sample every city represents another NUTS3 region and therefore the distance to the NUTS3 centroid is of special interest as it describes the distance to the respective city center. The socioeconomic variables, purchasing power per household and the number of households at the ZIP-code level, are extracted from the GfK-database, the population density per sqkm in a ZIP code area is calculated in ArcGIS. The last source is Reuters providing the 10-year interest rate for housing loans as a macro variable. The variables, their units and sources can be found in table 1.

Table 2 shows the heterogeneity within the largest seven German cities by displaying the variation of the variables describing the respective city. The heterogeneity between the cities emerges by comparing the variation of the individual variables across the cities. The appeal of using the quantile regression to explain time on market becomes apparent by investigating the variation in the variable of interest. A relative standard deviation ranging from 1.36 to 1.52 implies, that on average, the time a dwelling is advertised on the market deviates from the mean by 1.36 to 1.52 times the mean. In absolute values, the time on market within the cities deviates on average between 5.47 to 9.46 weeks from the mean. Across the cities, the variation in time on market rises along the distribution curve. While the average time a dwelling is placed on the market stretches from 3.75 to 6.71 weeks, the variation becomes more apparent along the distribution, with a spread from 15.6 weeks in Munich to 25.9 weeks in Dusseldorf for the 95th percentile. The presence of strong variation is not only true for the dependent variable, but also for the many covariates. The monthly rent in € per sqm ranges from a mean of 8.79 (95th-percentile: 13.90) in Berlin to 15.18 (95th-percentile: 21.67) in Munich, which are at the same time the locations with the lowest and highest mean in purchasing power. When observing the average building age, the relation is reversed. While on average the oldest dwellings are in Berlin, the mean and median in building age is lowest in Munich, indicating high development activity in the more recent past. Living area spans from an average of 70.89 sqm in Berlin to 79.01 sqm in Stuttgart. Within the cities, the standard deviation ranges from 29.21 sqm in Hamburg to 36.39

sqm in Munich. Although there have been many articles on the tremendous increase in the rent level of Berlin, the measure suggest the lowest overpricing based on hedonic characteristics among the largest seven German cities. This finding might support the idea that the rent level of Berlin is now able to recover, after the investment and maintenance backlog subsequent to the years of separation have been overcome. It is also interesting to note, that although on average the most households per ZIP code area are to be found in Cologne, the city has the second lowest average population density.

As the aim of the study is to investigate the impact of changes in the explanatory variables on the time on market of rental dwellings segmented by their liquidity level, it is of particular interest to identify, whether there are patterns in the dwelling characteristics, which can be used to explain the affiliation to the respective liquidity quantile. The analysis shows, that across all seven cities, the dwellings in the most liquid quantile, which is the 0.2-quantile, are on average the least expensive, the smallest, the oldest, have the least number of rooms, are located in ZIP codes with the least purchasing power, are closest to the city center, have the least dwellings offered for first occupancy, are the least renovated, are less overpriced, and are located in more densely populated ZIP codes. The dwellings assigned to the 0.8-quantile display on average 7.17% higher rent, are 25.72% larger, are 13.49% younger, have 15.33% more rooms, are located in ZIP codes with 3.16% higher purchasing power, are located 6.41% farer from the city center, have 70.61% more dwellings advertised for first occupancy, are 12.85% more renovated, are 47.70% more overpriced and are located in ZIP codes that are 12.07% less densely populated. Only in Munich and Frankfurt, the dwellings in the 0.8-quantile are 11.36% and 7.42% less renovated than those in the most liquid quantile.

Table 1: Variables and sources

Variable	**Unit**	**Source**			
		Empirica	GfK	ArcGIS	Reuters
Asking rent	€/m^2/p.m.	√			
Time on market	Weeks	√			
Living area	m^2	√			
Age	Years	√			
Rooms	Number	√			
With bathtub	Binary	√			
With built-in kitchen	Binary	√			
With parking slot	Binary	√			
With terrace	Binary	√			
With balcony	Binary	√			
With elevator	Binary	√			
Newly built dwelling	Binary	√			
Refurbished dwelling	Binary	√			
Degree of overpricing	Centered	√			
Gaussian longitude	Coordinate	√			
Gaussian latitude	Coordinate	√			
Distance to ZIP centroid	Km.			√	
Distance to NUTS3 centroid	Km.			√	
Households in ZIP	HHs/ZIP		√		
Purchasing power of HHs in ZIP	€/ HH /p.a.		√		
Population density	Persons/km^2/ZIP			√	
IR for housing loan 10 years	Effective interest rate in %				√
N		394,366			

Table 2: Descriptive statistics

Panel A: Berlin, ownership rate: 15.61%

Variable	**Mean**	**Median**	**StdDev**	**Min**	**Max**
Asking rent	8.79	8.29	2.61	1.19	30.41
Time on market	6.21	2.6	9.46	0	57.1
Living area	73.44	66.78	32.8	8	527.44
Age	57.11	51	35.63	-1	116
Degree of overpricing	0.15	0.11	0.96	-10.22	5.23
Households in ZIP	12,033.62	12,037	3613.12	98	20,434
Purchasing power of HHs in ZIP	35,446.42	34,437.95	5090.8	27,414.97	55,978.22
Population density	8076.04	7012.8	5330.85	86.34	17,700.64
N = 140,244					

Panel B: Hamburg, ownership rate: 24.14%

Variable	Mean	Median	StdDev	Min	Max
Asking rent	10.67	10.22	2.82	1.07	33.61
Time on market	5.01	2.2	7.3	0	40.5
Living area	70.89	65.53	29.21	10	530
Age	47.52	49	30.9	-1	116
Degree of overpricing	0.58	0.62	1.01	-10.9	5.58
Households in ZIP	11,122.58	10797	3410.62	469	17,979
Purchasing power of HHs in ZIP	43,584.38	42,779.87	7724.66	32,723.47	71,048.54
Population density	5813.02	5022.35	4160.61	207.84	17,700.64
N = 83,918					

Panel C: Munich, ownership rate: 25.23%

Variable	Mean	Median	StdDev	Min	Max
Asking rent	15.18	14.62	3.45	1.37	45
Time on market	3.79	1.7	5.69	0	33
Living area	76.59	71	36.39	8	435
Age	36.99	36	28.03	-1	116
Degree of overpricing	0.74	0.75	0.95	-9.88	5.49
Households in ZIP	11,525.96	12,327	3160.52	1842	16,896
Purchasing power of HHs in ZIP	55,888.76	54,060.67	6220.05	41,405.4	69,752.31
Population density	7127.11	5625.76	4453.88	696.21	17,700.64
N = 38,892					

Panel D: Cologne, ownership rate: 27.42%

Variable	Mean	Median	StdDev	Min	Max
Asking rent	9.93	9.73	2.22	2.16	32.5
Time on market	5.09	2.5	7.02	0	39.58
Living area	71.89	68.05	29.62	11.2	467.89
Age	42.73	45	26.67	-2	116
Degree of overpricing	0.86	0.89	0.88	-6.74	5.15
Households in ZIP	13,452.42	13,521	3548.82	3142	20,561
Purchasing power of HHs in ZIP	45,543.52	45,076.29	5760.97	34,685.48	58,827.02
Population density	4905.68	3571.58	3663.84	738.05	13,542.82
N = 42,022					

Panel E: Frankfurt, ownership rate: 20.67%

Variable	Mean	Median	StdDev	Min	Max
Asking rent	12.4	12.11	2.94	2.5	45.45
Time on market	5.95	2.8	8.23	0	45.79
Living area	77.9	72	36.07	10	600
Age	41.47	43	33.52	0	116
Degree of overpricing	0.9	0.88	0.94	-5.89	7.3
Households in ZIP	11,236.56	11,669	4282.48	7	20,945
Purchasing power of HHs in ZIP	47,495.1	46,692.74	6862.3	36,766.34	76,088.03
Population density	5536.71	4788.51	4172.71	0.53	17,700.64
N = 33,512					

Panel F: Stuttgart, ownership rate: 32.92%

Variable	Mean	Median	StdDev	Min	Max
Asking rent	11.24	10.88	2.62	2.08	28
Time on market	3.75	1.7	5.47	0	30.66
Living area	79.01	74	33.86	11	281
Age	46.1	47	31.56	0	116
Degree of overpricing	1.1	1.11	0.93	-7.04	4.91
Households in ZIP	10,463.69	10,943	3126.51	1104	15,899
Purchasing power of HHs in ZIP	47,261.47	46,393.16	4444.36	40,041.62	61,972.64
Population density	4759.66	3674.78	4160.55	876.69	17,700.64
N = 14,940					

Panel G: Dusseldorf, ownership rate: 24.08%

Variable	Mean	Median	StdDev	Min	Max
Asking rent	9.72	9.4	2.21	1.46	28
Time on market	6.71	3.3	9.1	0	50.16
Living area	75.52	70	33.69	10	520
Age	50.51	53	28.03	0	116
Degree of overpricing	0.66	0.66	0.87	-9.15	4.8
Households in ZIP	9744.15	9703	3001.67	335	15,045
Purchasing power of HHs in ZIP	47,917.18	46,140.99	5870.24	40,382.03	67,277.93
Population density	6387.52	5666.47	4537.08	599.62	17,700.64
N = 40,838					

Table 3: Average ToM in weeks per quantile

	0.2-quantile	0.3-quantile	0.4-quantile	Median	0.6-quantile	0.7-quantile	0.8-quantile
Berlin	0.54	1.00	1.52	2.24	3.26	4.82	7.26
Hamburg	0.44	0.74	1.20	1.84	2.69	4.02	6.04
Munich	0.30	0.61	1.00	1.48	2.14	3.04	4.46
Cologne	0.54	1.01	1.49	2.15	3.02	4.24	6.12
Frankfurt	0.59	1.10	1.66	2.41	3.49	5.00	7.34
Stuttgart	0.30	0.60	0.94	1.42	2.10	3.02	4.62
Dusseldorf	0.70	1.25	1.96	2.81	4.04	5.60	8.22

4. Estimation results

4.1. Results of the Cox survival regressions

In a first step, covariates boosting or limiting the time on market of rental dwellings on the German housing market as a whole have been considered. Therefore, a semiparametric proportional hazards model (PHM) has been estimated

according to

$$h(t_{ijkp}) = exp(X_{ip}\beta + Z_j\alpha + R_p\lambda + \mu_{ip}\delta_p + \mu_{ij}\rho_j) + e_{ijkp}. \tag{19}$$

Building upon Cajias and Freudenreich (2018), an improved specification with slightly different regressors has been estimated. The hazard function h of the time on market t depends on the matrix X containing the hedonic characteristics of dwelling i at observation period p, Z including time-invariant socioeconomic data on ZIP-code level j, R representing the time-varying effective 10-year interest rate for housing loans and μ_{ip} and μ_{ij} accounting for p time- and j spatial effects respectively. Building upon Cajias and Freudenreich (2017), further specifications with different regressors have been estimated leading to the set of regressors depicted in table 2. Time fixed effects as quarterly dummies ranging from 1 to 18 have been included to absorb seasonal fluctuations and to account for structural breaks. The construction dummies describe the period of time a dwelling was built in on a ten year range from 1910 till today. The results of the Cox survival regressions are presented in table 2. 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 range from 64.3% to 67.5%. Those values are common in survival studies and cannot directly be interpreted as or compared to the usual R^2.

For the estimations we do not use the asking rent in €per sqm, but the total asking rent of the respective dwelling in €in order to isolate the effect of the living area. The results show that an increase in the asking rent leads to a longer time on market across all seven cities. Cajias and Freudenreich (2018) were able to find the same positive relationship between rent and time on market for six out of the top seven German cities. This result is not very surprising, as an increase in asking rent is expected to have a negative implication on the liquidity of dwellings. Since the densely populated cities regularly show 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 curve. Asking rent, which is higher than the rent "justified" by the hedonic characteristics of the dwelling, surprisingly shortens time on market in all cities but Berlin. Cajias and Freudenreich (2018) found this liquidity enhancing relationship only for the two wealthiest cities within their sample, hence again a segmentation into liquidity submarkets might be useful to better detect market patterns and see whether this unexpected impact of overpricing is true for all dwellings in those cities. An increase in living area increases liquidity in five out of seven cities, whereas the number of rooms in a dwelling has the same positive effect on liquidity for all seven top-cities. Sur-

prisingly, the marketing time is shorter the older the dwelling and is longer for newly built and refurbished ones. The spatial gravity variables have a similar impact on time on market as in Cajias and Freudenreich (2018). An increase in the distance to the NUTS3 centroid, which is used as a proxy for the city center of those wide spanning cities, is associated with longer time on market. The distance to the ZIP code centroid again shows no statistically significant impact. The coefficients of the socioeconomic factors show the expected signs with purchasing power and population density reducing time on market. The effective 10-year interest rate for housing loans was included to account for market interactions between the property and the rental market. As only for two out of seven cities, a weak significant impact on time on market was found, it seems like interactions between the markets are negligible for the non-segmented sample.

4.2. Results of the censored quantile regression

In a second step, we use the same regressors as for the Cox survival regressions and estimate censored quantile regressions in order to get deeper insights into liquidity patterns. Therefore, for each TOP7 German city the rental market was divided into seven time on market quantiles. The results for the covariates of interest are shown in figure 2 to 5. Each plot shows the development of one coefficient β^{τ} over the liquidity quantiles τ. The main effects divided into quarterly factors, hedonic characteristics, socioeconomic characteristics and spatial gravity variables are reported in the following. It can be observed that the effect of all explanatory variables on the time on market increases with decreasing liquidity. That means, that for very liquid dwellings, a change in each of the explanatory variables has a relatively weak impact on the time on market compared to very illiquid dwellings. This is not surprising, since the descriptive statistics in table 3, show that the level as well as the range of the time on market is very low for the liquid quantiles and accelerates exponentially with rising iliquidity.

4.2.1. Quarterly time effects

The considered period, the period between the first quarter of 2013 and the second quarter of 2017, is characterized by low interest rates, high migration to Germany, especially to the metropolises and additionally way to little housing supply in these cities. As a consequence, vacancy has mostly been diminishing and real estate prices as well as rents have been increasing massively. Despite rising construction activity, building completion was insufficient to meet demand, leading to excess demand. Time fixed effects as quarterly dummies

Table 4: Results of Cox survival regressions

Hazard ratios and robust standard errors	Berlin	Hamburg	Munich	Cologne	Frankfurt	Stuttgart	Dusseldorf
Hedonic covariates - metric							
Log asking rent in €	0.448 0.071***	0.282 0.084***	0.113 0.116***	0.256 0.114***	0.189*** 0.115	0.123 0.188 ***	0.456 0.104***
Log living area	0.886 0.070	1.344 0.078***	2.934 0.105***	1.511 0.103***	1.819 0.103***	2.766 0.174***	0.954 0.099
Number of rooms	1.172 0.006***	1.178 0.008***	1.140 0.011***	1.139 0.011***	1.177 0.012***	1.120 0.016***	1.153 0.011***
Age	1.011 0.001***	1.005 0.002**	1.005 0.002*	0.998 0.002	1.014 0.002***	1.005 0.004	1.001 0.002
Degree of overpricing	0.910 0.015***	1.005 0.018	1.313 0.023***	1.103 0.024***	1.140 0.024***	1.271 0.040***	0.975 0.022
Hedonic covariates - binary							
With bathtub	0.874 0.006***	0.972 0.009**	0.960 0.012***	0.972 0.011*	0.900 0.013***	0.936 0.019***	0.900 0.011**
With built-in kitchen	1.043 0.010***	0.961 0.013**	1.181 0.018***	1.183 0.017***	1.238 0.018***	1.200 0.028***	1.144 0.016***
With parking slot	0.924 0.008***	0.899 0.010***	0.961 0.016***	0.999 0.014	0.955 0.017**	1.031 0.026	0.992 0.015
With terrace	0.928 0.009***	0.994 0.011	0.990 0.015	0.981 0.016	0.956 0.018*	1.050 0.024*	0.965 0.016*
With balcony	0.966 0.007***	1.031 0.010**	1.026 0.013*	1.025 0.013*	1.012 0.014	1.040 0.022	1.010 0.013
With lift	0.931 0.009***	0.899 0.012***	1.134 0.016***	0.882 0.015***	0.997 0.019	1.004 0.030	0.929 0.015***
Newly built	0.750 0.013***	0.937 0.017***	0.997 0.022	1.043 0.024	1.047 0.024	1.063 0.039	0.925 0.022***
Refurbished	0.980 0.001**	0.899 0.012***	1.008 0.015	0.997 0.012	0.972 0.015	0.981 0.023	0.925 0.012***
Spatial gravity variables							
Longitude	0.702 0.031***	0.522 0.045***	0.831 0.105	0.262 0.117***	1.444 0.145*	0.319 0.228***	0.102 0.177***
Latitude	0.533 0.061***	1.082 0.085	1.385 0.207	0.324 0.181***	0.422Â´ 0.309**	1.418 0.294	0.162 0.267***
Distance to ZIP centroid	1.018 0.008*	1.040 0.010***	0.975 0.019	1.004 0.013	1.027 0.019	0.944 0.024*	0.990 0.017
Distance to NUTS3 centroid	0.956 0.001***	0.958 0.002***	0.983 0.004***	0.955 0.004***	0.977 0.005***	0.967 0.007***	0.980 0.003***
Time-invariant effect at ZIP level							
Log purchasing power	1.490 0.043***	1.672 0.056***	2.200 0.097***	1.796 0.086***	1.713 0.096***	1.436 0.154*	1.744 0.089***
Log number of households	0.930 0.012***	0.852 0.016***	1.081 0.026**	0.936 0.024**	1.043 0.020*	1.139 0.029***	1.029 0.019
Log population density	1.066 0.006***	1.212 0.008***	1.125 0.015***	1.131 0.012***	1.162 0.016***	1.039 0.018*	1.071 0.011***
Financial conditions at day of release							
Effective 10Y interest rate for housing loan	1.141 0.057*	1.069 0.074	0.967 0.111	1.121 0.101	0.957 0.114	0.889 0.169	1.198 0.102
Fixed effects							
Construction dummies	Included						
Quarterly dummies	Included						
Intercept	Included						
Spatial adjustment of standard errors via Win-Lei	Considered						
R^2-concordance	65.1%	67.0%	67.5%	64.7%	65.4%	66.9%	64.3%
N	140,244	83,918	38,892	42,022	33,512	14,940	40,838

have been included to capture the time trend of the time on market. This time trend can be observed in figure 3 for quantiles representing high, medium and low liquidity. The base quarter is the first quarter of 2013, thus all changes are with respect to this basis. Comparing the highly liquid, average liquid and highly illiquid segments, the change in the time on market is greatest for highly illiquid dwellings, i.e. in the 0.8-quantile. This is not surprising as the range in the time on market is relatively large for the greater quantiles compared to the smaller ones. It is visible, that at the beginning of the observation period, time on market has been increasing relative to the basis in almost every city across all liquidity quantiles. This might to some extent be due to the selected base quarter, which is the first quarter of the year and this is when time on the market is usually relatively low for rental dwellings.

At the median and the 0.8-quantile, the time on market started decreasing in 2014 and kept this direction until the end of the observation period across all seven cities.

For highly liquid dwellings, the evolution of the time on market shows slight differences between the cities. Besides Hamburg, all cities exhibit a declining development of the time on market starting in the year 2014.

The increasing time on the market compared to the base quarter at the beginning of the period under consideration was by far strongest in Dusseldorf across all quantiles. This might be due to the strong increase in construction completions in 2014 compared to 2013, however, with construction completions still being on a very low level. Also, the increase in population has been modest in Dusseldorf compared to the other top seven German cities. The decline in the time on market from about 2014 onwards has been particularly strong in Cologne, Frankfurt and Dusseldorf across all liquidity quantiles. A reason for that might be that the rent level has been increasing only moderately in Dusseldorf, Cologne and Frankfurt relative to the other cities considered. Furthermore, in Frankfurt population and hence the demand for space have been increasing extensively, leading to more rental growth in the latest periods. In Berlin, a huge decrease in the time on market can be observed at the median and especially for very iliquid dwellings. Again, this might be due to the relatively flat rent development as well as its still very low level. Moreover, the number of migrants clearly exceeds the level of construction completions and new building land is scarce. For highly liquid dwellings in Berlin, however, time on market has been decreasing only moderately compared to the other cities. The reduction in time on market in Hamburg is situated in the middle when comparing cities and across quantiles accelerates with illiquidity. This moderate development can be explained by a very active construction activity in Hamburg, resulting in only very little rental growth. The cities with very small changes in the time on

market across all quantiles are Munich and Stuttgart. These cities exhibit the largest rental increase. Especially in Munich, the number of inhabitants has been vastly rising, so that an enormous excess demand for rental dwellings has been emerging along with vacancy rates close to zero. Furthermore, purchasing power per person is highest in Munich and unemployment is lowest. Thus, the time on market development might seem somehow surprising. However, with a glance at the descriptive statistics in table 3, it might to some extent be due to the anyway relatively low levels of time on market in each of the respective quantiles compared to the other cities.

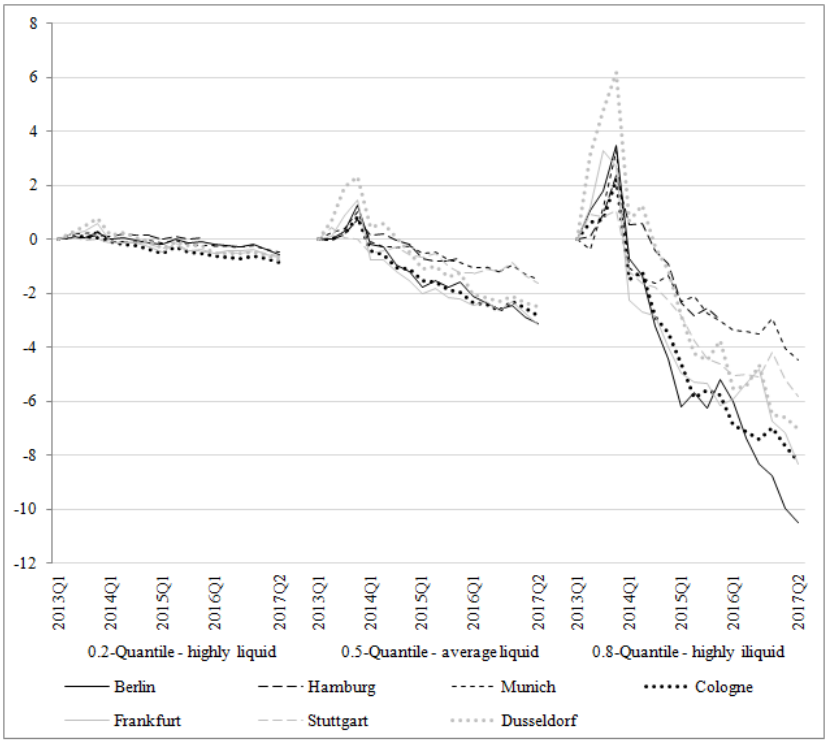

Figure 2: Estimated quantile regression coefficients β_{ik}^{τ} of the quarterly dummies

4.2.2. Hedonic characteristics

It is possible to cluster the hedonic variables in three groups based on their impact on time on market. While asking rent unsurprisingly 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, living area and the degree of overpricing show the opposite impact for all statistically significant quantiles across the seven cities. The building age, first occupancy, and renovation show distinct effects, which are dependent on the location and the liquidity level of

the dwelling.

For the covariate asking rent, the magnitude of the impact on time on market rises exponentially with the level of illiquidity for all cities. While the impact of an increase in asking rent across the liquid quantiles is weakest for dwellings located in Berlin, the weakest impact across the more illiquid quantiles is found for properties in Cologne and Stuttgart respectively. Within the most illiquid quantile, the impact of an increase in asking rent on time on market converges for Stuttgart, Cologne, and Hamburg. The strongest impact for the highly liquid quantiles is found for Munich but changes to Frankfurt with the 40% percentile. The spread of the impact widens with increasing illiquidity. While a ten percent increase in asking rent within the most liquid quantile results in 6.2% to 14.9% higher time on market in Berlin and Munich, the same increase in rent lengthens time on market from 115.5% to 196.8% in Stuttgart and Frankfurt for the most illiquid quantile.

The impact of an increase in living area shows the expected opposite pattern, as larger living area is a positive factor for the marketability of property, all else equal. The weakest impact of an increase in living area on time on market is observed for dwellings in Berlin and changes only for the most illiquid quantile where the impact is lower for dwellings in Hamburg and Cologne. The strongest impact of an increase in living area is found for dwellings in Munich up to the 0.6-quantile. Afterwards the strongest impact on time on market is found for dwellings in Frankfurt. Like for the variable asking rent, the variety of the impact on time on market across the cities rises along the distribution. For the most liquid dwellings, the impact of a ten percent increase in living area decreases time on market by 1.0% in Berlin to 7.6% in Munich. The same ten percent increase in living area decreases the time on market for the least liquid dwellings by 29.3% in Hamburg and by 41.9% in Frankfurt.

Unlike the impact of a change in asking rent and living area, the effect of a change in age does not have the uniform effect on time on market across all quantiles and cities, nor is it statistically significant for all of them. The impact is very inconclusive and insignificant for Munich, Cologne, Stuttgart and Dusseldorf. For Frankfurt and Berlin, the negative implication of an increase in age on time on market lies within a similar range. In Frankfurt and Berlin, an increase by ten years shortens the time on market for the most liquid dwellings by 8.2% and 4.7%. Again, the magnitude of the effect expands with rising illiquidity and results in a 71.4% to 70.4% shorter marketing time for the two cities.

For the degree of overpricing, the results are very conclusive yet surprising. For all statistically significant quantiles a higher level of overpricing shortens the marketing time of dwellings. It is even more surprising, that for the any-

way less demanded properties, a higher overpricing shortens the marketing time even more, as the effect increases exponentially along the distribution. Up to the 0.6-quantile, the impact of a change in overpricing on time on market is strongest for Munich. For the more illiquid quantiles the effect of a change in overpricing converges for Frankfurt and Munich. The weakest effect up to the 0.7-quantile is found for dwellings in Berlin, while for the most illiquid dwellings the impact on time on market is weaker in Hamburg.

Whether the landlord offers the dwelling for first occupancy has a quite distinguishable impact on the marketing time across the cities. As no changes in direction are observable, the impact on time on market is either consistently positive, like in Berlin, or consistently negative. While for Dusseldorf no statistically significant impact was found, the fact a dwelling is offered for first occupancy shows highly significant impact on the time on market of all dwellings in Berlin, Munich, Cologne, Frankfurt, and Stuttgart. In Hamburg, the impact is only statistically significant for the quantiles 0.3 to 0.6. Again, the strongest effect on time on market is observed for the highly illiquid quantiles. For Berlin, the positive thus time on market prolonging effect of whether the dwelling is offered for fist occupancy ranges from a 22.6% higher time on market for the most liquid quantile to a 594.1% higher time on market for the most illiquid quantile. The strongest negative impact for the more liquid quantiles is found in Stuttgart. For the more illiquid quantiles, the dwellings in Frankfurt with a 84.6% lower time on market show the highest decrease in marketing time when a dwelling is offered for first occupancy.

Whether the dwelling is offered on the market right after a renovation took place, has a consistently positive and statistically significant impact on time on market for dwellings in Hamburg and Dusseldorf. In Berlin, Frankfurt, and Cologne the impact on the marketing time changes from positive to negative along the distribution and temporarily loses the statistically significance. For dwellings located in Munich and Stuttgart, the fact a dwelling is advertised as renovated, has a negative impact on time on market, irrespective of the liquidity level but is only statistically significant across all quantiles for Munich. As for the other explanatory variables, the spread of the variable's impact on time on market widens with increasing illiquidity. While the weakest positive impact on time on market for the most liquid dwellings in Berlin increases the marketing time by 5.4% if a dwelling is advertised as renovated, the impact on the marketing time for the most illiquid dwellings in Hamburg results in a 67.4% longer time on market. For dwellings in Munich, the negative impact on time on market increases from 6.9% to 42.8%.

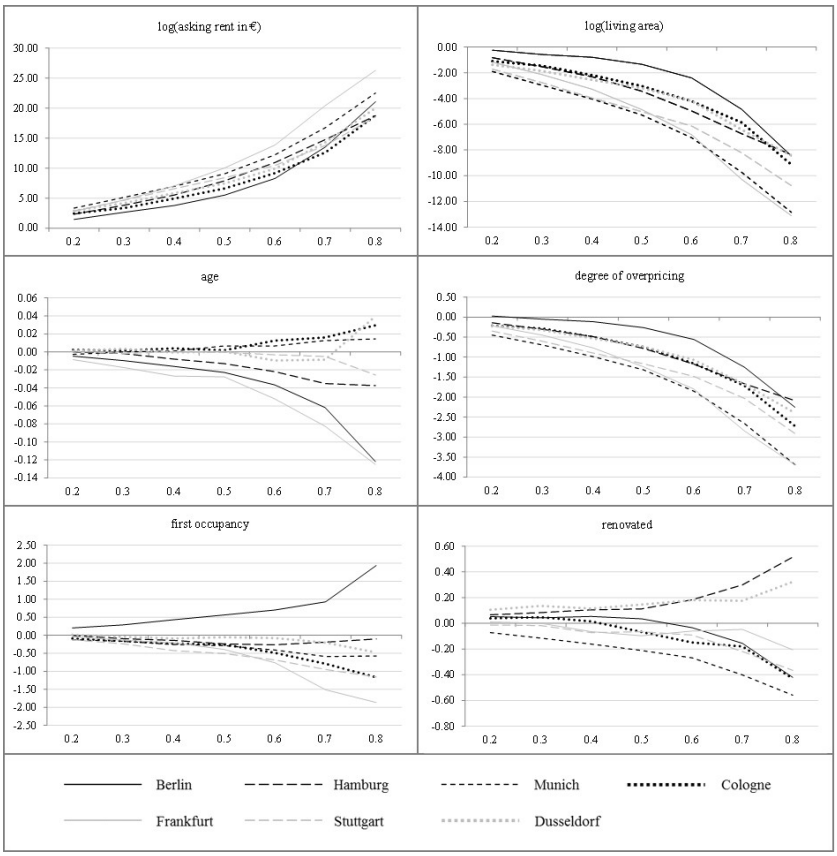

Figure 3: Estimated quantile regression coefficients β_{ik}^{τ} of the hedonic variables

4.2.3. Spatial gravity variables

Considering the spatial variables, the distance to the NUTS3 center is of particular interest, since the study analyses the largest seven German cities geographically restricted by its NUTS3 boundaries. Therefore, the distance from an individual dwelling to the center of the NUTS3 region can be interpreted as its distance to the geographical city center.

Not surprisingly, a higher distance to the center is extending the time a rental dwelling is listed on the market. The effect is getting stronger for the highly illiquid quantiles. While for the more liquid half of the quantiles, dwellings in Cologne are affected the most, a higher distance to the city center has a larger impact on the marketing time of dwellings in Berlin for the more illiquid quantiles. With 8.44 kilometers to the centroid of the NUTS3 region, the rental dwellings in Berlin display on average the highest distance to the approximated city center. The dwellings assigned to the most illiquid quantile are on average

6.91% further from the city center as the dwellings in the 0.2-quantile. The weakest impact of an increase in distance is found for dwellings in Munich and is independent of the liquidity level.

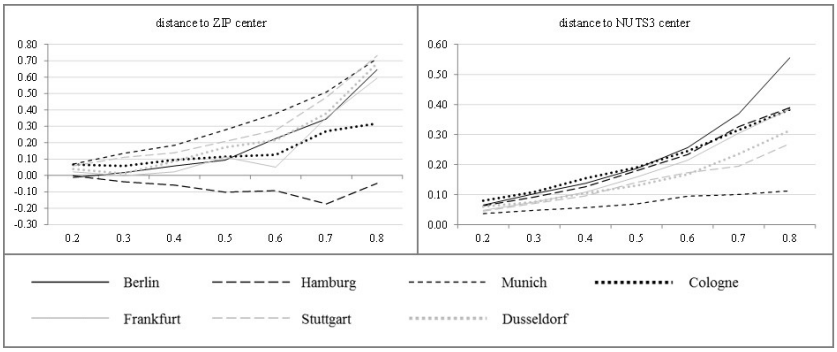

Figure 4: Estimated quantile regression coefficients β_{ik}^{τ} of the spatial gravity variables

4.2.4. Socioeconomic characteristics

The socioeconomic factors purchasing power, number of households per ZIP and population density mostly exhibit a negative impact on time on market. This result is generally common across all cities and all liquidity quantiles. The impact of the socioeconomic factors, hence, can be summarized as a demand effect. This is not surprising, as massive excess demand is visible in the largest German cities.

An increase in purchasing power leads to more demand on the real estate market and thus decreases the time a dwelling stays on the market. The effect rises from very liquid to highly illiquid segments. While this development is modest for the relatively liquid half of the quantiles, it accelerates in the illiquid half, meaning that purchasing power does not affect liquid dwellings much but has a huge impact on dwellings that stay on the market for a longer period. For dwellings located in Stuttgart, an increase in purchasing power has the weakest effect on time on market across all cities. The impact of a 10% increase evolves from a 0.9% shorter time on market for the most liquid dwellings to a 15.6% shorter time on market for the most illiquid ones. The impact of purchasing power across the more liquid half of the distribution is strongest in Munich but changes to Dusseldorf and Frankfurt with rising illiquidity. While a 10% increase in purchasing power shortens the marketing time of the most liquidid dwellings in Munich by 5.2%, the same increase shortens the marketing time of very illiquid dwellings in Dusseldorf by 39.1%. Comparing the most liquid with the

most illiquid quantile in Dusseldorf, a 10% increase in purchasing power results in a 4.1% lower time on market, compared to 39.1%. Consequntly, the effect accelerates with growing illiquidity. However, high illiquidity can be attributed to higher levels of purchasing power. This might indicate that richer households spend more time for the search and matching process.

The effect of the number of households in a ZIP code area on time on market shows a similar picture. With decreasing liquidity, the effect of an additional household in the ZIP code area increases for five out of seven cities. In Cologne and Berlin, the effect is mainly insignificant. Across the highly illiquid quantiles, the impact of an increase in households on time on market is strongest for the relatively small cities Dusseldorf, Stuttgart and Frankfurt. In those cities, a 10% increase in the number of households per ZIP code area shortens the marketing time by 6.5% in Dusseldorf to 7.1% in Frankfurt. In the larger cities Munich and Berlin, the impact is of smaller magnitude and results in a 5.5% to 2.5% shorter time on market. Only for Hamburg, an increase in the number of households lengthens the marketing time. The impact increases from 0.7% for the highly liquid dwellings to 1.0% for the highly illiquid dwellings and loses its significance for the 0.6- and 0.7-quantile.

While for dwellings located in Hamburg, an increase in the number of households per ZIP code area has a time on market prolonging effect, an increase in the population density within the ZIP code area shows a strong positive effect on the liquidity of dwellings. The marketing time of dwellings in Stuttgart on the other hand was affected the most by an increase in the number of households but is affected the least by an increase in population density. For the other cities, the impact of an increase in population density is relatively uniform and increases linearly with the level of illiquidity. Even for the most illiquid quantile the impact of a 10% increase in population density on time on market ranges only from 3.1% to 3.9%. For the most illiquid dwellings in Stuttgart the same percentage increase in population density results in a 1.1% shorter marketing time, while for Hamburg the decrease in marketing time is 7.5%.

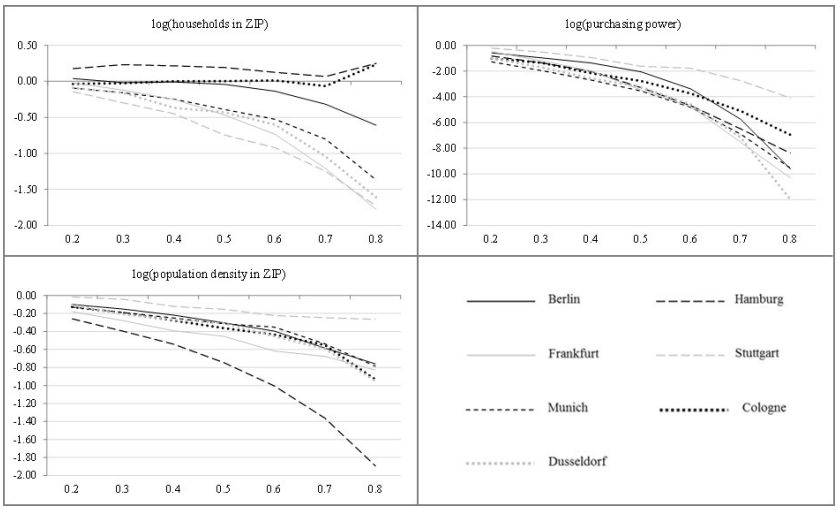

Figure 5: Estimated quantile regression coefficients β_{ik}^{τ} of the socioeconomic variables

5. Conclusion

To the best of our knowledge, this study is the first to apply censored quantile regressions for real estate data, in order to estimate the determinants of time on market for rental dwellings. Using an extensive data set covering 394,366 observations of rental transactions within the seven largest German cities, the study is able to identify the factors that drive time on market in low-, medium and high-liquidity market segments. For almost all regression coefficients we find consistent signs across all quantiles of the time on market distribution, i.e. the proportional hazard assumption, underlying the Cox model, is not violated. While the impact of a change in the explanatory variables may differ in magnitude across the liquidity quantiles and cities, it is quite distinct that the very illiquid dwellings are affected much more by those changes than the more liquid dwellings. Thus, we prove that segmenting the market into different liquidity segments is a crucial point for the liquidity analysis of rental dwellings on the real estate market.

References

Allen, Marcus T., Ronald C. Rutherford, and Thomas A. Thomson (2009) "Residential asking rents and time on the market," *The Journal of Real Estate Finance and Economics*, Vol. 38, pp. 351–365.

Anglin, Paul M., Ronald Rutherford, and Thomas M. Springer (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*, Vol. 26, pp. 95–111.

Bassett Jr., Gilbert W. and Hsiu-Lang Chen (2001) "Portfolio style: Return-based attribution using quantile regression," *Empirical Economics*, Vol. 26, pp. 293–305.

Belkin, Jacob, Donald J. Hempel, and Dennis W. McLeavey (1976) "An empirical study of time on market using multidimensional segmentation of housing markets," *Real Estate Economics*, Vol. 4, pp. 57–75.

Beyerlein, Andreas, André M. Toschke, and Rüdiger von Kries (2008) "Breastfeeding and childhood obesity: shift of the entire BMI distribution or only the upper parts?" *Obesity*, Vol. 16, pp. 2730–2733.

Buchinsky, Moshe (1994) "Changes in the US wage structure 1963-1987: Application of quantile regression," *Econometrica: Journal of the Econometric Society*, pp. 405–458.

——— (1995) "Quantile regression, Box-Cox transformation model, and the US wage structure, 1963–1987," *Journal of Econometrics*, Vol. 65, pp. 109–154.

Cajias, Marcelo and Philipp Freudenreich (2018) "Exploring the determinants of liquidity with big data - market heterogeneity in German markets," *Journal of Property Investment and Finance*, Vol. 36, pp. 1–17.

Chaudhuri, Probal, Kjell Doksum, Alexander Samarov et al. (1997) "On average derivative quantile regression," *The Annals of Statistics*, Vol. 25, pp. 715–744.

Cole, Timothy J. and Pamela J. Green (1992) "Smoothing reference centile curves: the LMS method and penalized likelihood," *Statistics in Medicine*, Vol. 11, pp. 1305–1319.

Cox, David R. (1972) "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 34, pp. 187–220.

Crowley, John and Marie Hu (1977) "Covariance analysis of heart transplant survival data," *Journal of the American Statistical Association*, Vol. 72, pp. 27–36.

Doksum, Kjell (1974) "Empirical probability plots and statistical inference for nonlinear models in the two-sample case," *The Annals of Statistics*, pp. 267–277.

Doksum, Kjell A. and Miriam Gasko (1990) "On a correspondence between models in binary regression analysis and in survival analysis," *International Statistical Review/Revue Internationale de Statistique*, pp. 243–252.

Farmer, Michael and Clifford Lipscomb (2010) "Using quantile regression in hedonic analysis to reveal submarket competition," *Journal of Real Estate Research*, Vol. 32, pp. 435–460.

Fitzenberger, Bernd and Ralf A. Wilke (2010) "New insights on unemployment duration and post unemployment earnings in Germany," *Oxford Bulletin of Economics and Statistics*, Vol. 72, pp. 794–826.

Goodman, Allen C. and Thomas G. Thibodeau (2007) "The spatial proximity of metropolitan area housing submarkets," *Real Estate Economics*, Vol. 35, pp. 209–232.

Guasch, J. Luis and Robert C. Marshall (1985) "An analysis of vacancy patterns in the rental housing market," *Journal of Urban Economics*, Vol. 17, pp. 208–229.

Harder, Thomas, Renate Bergmann, Gerd Kallischnigg, and Andreas Plagemann (2005) "Duration of breastfeeding and risk of overweight: a meta-analysis," *American Journal of Epidemiology*, Vol. 162, pp. 397–403.

Hendricks, Wallace and Roger Koenker (1992) "Hierarchical spline models for conditional quantiles and the demand for electricity," *Journal of the American Statistical Association*, Vol. 87, pp. 58–68.

Horowitz, Joel L. and George R. Neumann (1987) "Semiparametric estimation of employment duration models," *Econometric Reviews*, Vol. 6, pp. 5–40.

Kalra, Rajiv and Kam C. Chan (1994) "Censored sample bias, macroeconomic factors, and time on market of residential housing," *Journal of Real Estate Research*, Vol. 9, pp. 253–262.

Kang, Han and Mona Gardner (1989) "Selling price and marketing time in the residential real estate market," *Journal of Real Estate Research*, Vol. 4, pp. 21–35.

Koenker, Roger (2008) "Censored quantile regression redux," *Journal of Statistical Software*, Vol. 27, pp. 1–25.

Koenker, Roger and Gilbert Bassett Jr. (1978) "Regression quantiles," *Econometrica: Journal of the Econometric Society*, pp. 33–50.

Koenker, Roger and Yannis Bilias (2002) "Quantile regression for duration data: a reappraisal of the Pennsylvania reemployment bonus experiments," in *Economic Applications of Quantile Regression*: Springer, pp. 199–220.

Koenker, Roger and Olga Geling (2001) "Reappraising medfly longevity: a quantile regression survival analysis," *Journal of the American Statistical Association*, Vol. 96, pp. 458–468.

Lehmann, Erich L. and H.J. D'abrera (1975) *Nonparametrics: statistical methods based on ranks.*: Holden-Day.

Liao, Wen-Chi and Xizhu Wang (2012) "Hedonic house prices and spatial quantile regression," *Journal of Housing Economics*, Vol. 21, pp. 16–27.

Lüdemann, Elke, Ralf A Wilke, and Xuan Zhang (2006) "Censored quantile regressions and the length of unemployment periods in West Germany," *Empirical Economics*, Vol. 31, pp. 1003–1024.

Machado, José AF and José Mata (2000) "Box-Cox quantile regression and the distribution of firm sizes," *Journal of Applied Econometrics*, Vol. 15, pp. 253–274.

Mak, Stephen, Lennon Choy, and Winky Ho (2010) "Quantile regression estimates of Hong Kong real estate prices," *Urban Studies*, Vol. 47, pp. 2461–2472.

Malpezzi, Stephen (2003) "Hedonic pricing models: a selective and applied review," *Housing Economics and Public Policy*, pp. 67–89.

McGreal, Stanley, Alastair Adair, Louise Brown, and James R. Webb (2009) "Pricing and time on the market for residential properties in a major UK city," *Journal of Real Estate Research*, Vol. 31, pp. 201–234.

Ong, Seow Eng and Yen Ching Koh (2000) "Time on-market and price trade-offs in high-rise housing sub-markets," *Urban Studies*, Vol. 37, pp. 2057–2071.

Owen, Christopher G., Richard M. Martin, Peter H. Whincup, George D. Smith, and Derek G. Cook (2005) "Effect of infant feeding on the risk of obesity across the life course: a quantitative review of published evidence," *Pediatrics*, Vol. 115, pp. 1367–1377.

Pandey, Ganesh R. and Van-Thanh-Van Nguyen (1999) "A comparative study of regression based methods in regional flood frequency analysis," *Journal of Hydrology*, Vol. 225, pp. 92–101.

Peng, Limin and Yijian Huang (2008) "Survival analysis with quantile regression models," *Journal of the American Statistical Association*, Vol. 103, pp. 637–649.

Portnoy, Stephen (2003) "Censored regression quantiles," *Journal of the American Statistical Association*, Vol. 98, pp. 1001–1012.

Powell, James L. (1984) "Least absolute deviations estimation for the censored regression model," *Journal of Econometrics*, Vol. 25, pp. 303–325.

——— (1986) "Censored regression quantiles," *Journal of Econometrics*, Vol. 32, pp. 143–155.

Rose, Stephen J. (1992) *Social Stratification in the United States*: New Press, New York.

Royston, Patrick and Douglas G Altman (1994) "Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling," *Applied Statistics*, Vol. 43, pp. 429–467.

Schmillen, Achim and Joachim Möller (2010) "Determinants of lifetime unemployment: a micro data analysis with censored quantile regressions."

Stigler, Stephen M (1984) "Studies in the history of probability and statistics XL Boscovich, Simpson and a 1760 manuscript note on fitting a linear relation," *Biometrika*, Vol. 71, pp. 615–620.

Taylor, James W (1999) "A quantile regression approach to estimating the distribution of multiperiod returns," *The Journal of Derivatives*, Vol. 7, pp. 64–78.

Turnbull, Geoffrey K. and Jonathan Dombrow (2006) "Spatial competition and shopping externalities: Evidence from the housing market," *The Journal of Real Estate Finance and Economics*, Vol. 32, pp. 391–408.

Turnbull, Geoffrey K., Jonathan Dombrow, and Clemon F. Sirmans (2006) "Big house, little house: relative size and value," *Real Estate Economics*, Vol. 34, pp. 439–456.

Wehby, George L., Jeffrey C. Murray, Eduardo E. Castilla, Jorge S. Lopez-Camelo, and Robert L. Ohsfeldt (2009) "Prenatal care effectiveness and utilization in Brazil," *Health Policy and Planning*, Vol. 24, pp. 175–188.

Wei, Ying, Anneli Pere, Roger Koenker, and Xuming He (2006) "Quantile regression methods for reference growth charts," *Statistics in Medicine*, Vol. 25, pp. 1369–1382.

Wood, John H. and Norma L. Wood (1985) *Financial markets*: Harcourt Brace Jovanovich.

Yang, Song (1999) "Censored median regression using weighted empirical survival and hazard functions," *Journal of the American Statistical Association*, Vol. 94, pp. 137–145.

Yavas, Abdullah and Shiawee Yang (1995) "The strategic role of listing price in marketing real estate: theory and evidence," *Real Estate Economics*, Vol. 23, pp. 347–368.

Yu, Keming, Zudi Lu, and Julian Stander (2003) "Quantile regression: applications and current research areas," *Journal of the Royal Statistical Society: Series D (The Statistician)*, Vol. 52, pp. 331–350.

Zietz, Joachim, Emily Norman Zietz, and G. Stacy Sirmans (2008) "Determinants of house prices: a quantile regression approach," *The Journal of Real Estate Finance and Economics*, Vol. 37, pp. 317–333.