EXPLORING THE DETERMINANTS OF LIQUIDITY WITH BIG DATA – MARKET HETEROGENEITY IN GERMAN MARKETS

Marcelo CAJIAS [a]

Philipp FREUDENREICH [b]

[a] PATRIZIA Immobilien AG, Fuggerstraße 26, 86150 Augsburg, GERMANY

[b] University of Regensburg, IRE|BS Institute, Universitätsstraße 31, 93053 Regensburg, GERMANY

ABSTRACT

Purpose – The purpose of this paper is to examine the market liquidity (time-on-market) and its determinants for rental dwellings in the largest seven German cities with big data.

Design/methodology/approach – The determinants of time-on-market are estimated with the Cox proportional hazards model. Hedonic characteristics as well as socioeconomic and spatial variables are combined with different fixed effects and controls for non-linearity to maximise the explanatory power of the model.

Findings – Higher asking rent and larger living space decrease the liquidity on all seven markets, while dwelling's age, the number of rooms and proximity to the city centre fasten the letting process. For the linear and non-linear hedonic characteristics heterogeneous implications are found.

Practical implications – The findings are of interest for institutional and private landlords as well as governmental organizations in charge of housing and urban development.

Originality/value – It is the first paper to deal with liquidity of rental dwellings in the seven most populated cities of Europe's second largest rental market by applying the Cox proportional hazards model. Furthermore, the German rental market is of particular interest, as approximately 60% of all rental dwellings are owned by private landlords and the German market is organized polycentric.

Keywords Liquidity/ Time-on-market; Housing real estate; Big data; Cox proportional hazards model; Non-linearity

Paper type Research paper

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.

EXPLORING THE DETERMINANTS OF LIQUIDITY WITH BIG DATA – MARKET HETEROGENEITY IN GERMAN MARKETS

1. Introduction

Financial assets such as stocks and bonds are traded in tremendous volumes, turning over billions of dollars within seconds and without any spatial constraints. In contrast, the transaction process of direct real estate is more complex, often consuming several months. When it comes to residential real estate a matching may be even more difficult as they are strongly determined by individual preferences of homebuyers and expectations of homesellers. A general understanding of liquidity in direct real estate is therefore essential for market players, either private, institutional or governmental, not only in order to derive investment strategies, but to assess market fundamentals and cyclical movements as well as political implications. Moreover, the instruments for efficiently capturing the factors boosting and dampening liquidity are crucial and far from being trivial, as liquidity in terms of "time" requires advanced econometric modelling. To capture the uncertainty of finding a match as well as the time a property is advertised on the market, liquidity in residential real estate literature is widely proxied by time-on-market (TOM). In this context, this paper explores liquidity of direct real estate focussing on the seven biggest German rental housing markets by means of advanced semiparametric survival techniques. The aim of the study is to explore liquidity concepts and examining the factors that determine liquidity such as linear, binary, spatial as well as possible non-linear effects with big data, in order to derive similarities and of course divergences between the cities. The paper may serve as a guide for market players and policy makers when conducting liquidity analysis on and understanding future developments in rental housing markets.

Since their establishment within the real estate literature, survival models have been adapted by various researchers to estimate the determinants of time-on-market. Kluger and Miller (1990) introduced the semi-parametric Cox proportional hazards model based on Cox (1972) to real estate studies, which allows a more flexible application without any a priori assumptions regarding the distribution of the baseline hazard in contrast to the widely used Weibull model. Studies following this approach are Krainer (1999), Smith (2010), Hoeberichts et al. (2013), Cirman et al. (2015), among others.

In search for an instrument capturing "users' taste" for dwellings and its effect on liquidity, Haurin (1988) developed an atypicality index and shows that for more atypical dwellings, the distribution of offers is prone to a wider variation. A dwelling is defined as atypical when its hedonic properties deviate substantially from the mean hedonic market characteristics, e.g. a dwelling with 150 m², 1 room, in the 10th floor without elevator is rather atypical. Nowadays, atypicality is a widely recognised factor in hedonic survival regressions as seen in Krainer (1999), Anglin et al. (2003), Bourassa et al. (2009), Haurin et al. (2010; 2013) and Hoeberichts et al. (2013), among others.

The signalling effect by setting the initial list price is also a widely researched area. Glower et al. (1998) for example, began to investigate the impact of the percentage difference in observed list price from the expected list price. Anglin et al. (2003) extended this approach and introduced a new explanatory variable in the context of liquidity called degree of overpricing (DOP). They defined the variable as the percentage deviation of an individual property's list price from the empirically estimated market list price. While they found that abnormal list prices, i.e. overpricing, increase the marketing time of houses, further application can also be found in Hoeberichts et al. (2013) and Cirman et al. (2015).

Over the last years, more and more emphasize is being placed on spatial effects when modelling the price and time-on-market of residential properties. Many articles have tested the theory of market segmentation within residential real estate markets, concluding that the inclusion of spatial variables improves the explanatory power of real estate pricing models e.g. Goodman and Thibodeau (2007), Turnbull and Dombrow (2006), Pavlov (2000), Fik et al. (2003), Bourassa et al. (2010) and Cirman et al. (2015) among others. Smith (2010) was the first to specify a Cox-model containing school districts and Cartesian coordinates. He found that, while the school district dummies and the coordinates are by itself statistically significant and show a large impact on the liquidity, the combination of both yields the largest explanatory power.

To our knowledge, the first study to estimate time-on-market in residential rental markets was conducted by Allen et al. (2009). The authors focus on the Dallas-Fort Worth area with a sample of over 20,000 listings and more than 11,000 corresponding letting contracts. Using a Weibull hazard model, the authors conclude that after the reset of asking rent initially overpriced by 15%, the landlords face 9.5 days longer time-on-market on average. Due to the initial overpricing and thus longer TOM, those landlords also have to accept a contract rent which is on average 5.2% below the hedonically estimated rent. Cajias et al. (2015) in contrast used a similar approach to estimate the effect of energy consumption on time-on-market for the German rental market. Using a Cox-model, the authors calculated the

odds of a dwelling to be let, dependent on energy consumption and show that dwellings with higher energy consumption relative to the most energy efficient dwellings stayed on the market for a longer time.

This paper investigates the determinants of liquidity on one of the largest rental markets in Europe. In Germany, home-ownership (ca. 45%) is at such a low level because of a large stock of high quality subsidized social housing built after World War II, low tax benefits for owners and a rather liberal rental market (Voigtlaender, 2009). A profound understanding of liquidity is therefore not only relevant for institutional landlords, but for millions of private providers of living space, as of those roughly 14.5 million rental properties approximately 60% are owned and let by individuals. At the same time, Germany provides a unique research field because of the polycentric market organization. In comparison to other European countries such as England or France, Germany divides its political, social and economic functions in up to seven cities. Each of those seven cities developed its own field of specialization, for example Frankfurt being the financial capital, Stuttgart being seen as automobile city and Munich as a hydride between new technologies and beer production. Therefore, the dataset contains a socially, culturally and economically diversified overview of major urban areas all over Germany and the study is able to yield comprehensive results explaining liquidity for the German rental market in an urban context.

The paper is organized as follows. The second section describes the methodology employed for the study, whereas the third section describes the data. The fourth section presents the results.

2. Econometric research approach

Prior to deriving the model, some statistical elements in the estimation of survival regression are to define. The time period (T) a flat is offered on the market corresponds to a continuous positive response variable without zeros and is interpreted as the duration of an event (t), in our case the time in weeks, before the occurrence of the letting agreement. Two main measures are important for understanding and estimating survival models: the survival function (S) and the hazard rate function (h). Formally they are expressed as:

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

$$h(t) = \frac{P(t<T\leq t+\Delta t|T>t)}{\Delta t} \tag{2}$$

While the survival function yields the probability that a dwelling survives until a certain time t, the hazard specifies the rate of failure at $T = t$ given that the flat survived up to time t. Since the numerator in equation (2) corresponds to a conditional probability and the denominator is an elapse of time, the hazard function gives the probability or rate of

"mortality" per units of time. Very important in survival analysis is the fact that some observations or dwellings do not change their event status, either because they remain available on the market or the landlord does not change the status in the Multiple Listing Service (MLS) database. In this case, the response variable is said to be right-censored. While simple models such as Kaplan-Meier or Kernel estimators manage to estimate the survival function, they are unable to control for the censoring effect properly. Cox hazard models do account for censoring in the response variable as they transform the response into a count variable per unit of time. In other words, the proportional Cox hazard model decomposes the time of an event in units of time incorporating censoring into the count regression. Since the response variable is expressed as time, survival models estimate a conditional survival probability for an event for each observation rather than estimating a single fitted value in the sense of the traditional OLS regression.

The Cox hazard model explains the factors that boost or restrict the letting process of a dwelling as a probability function after controlling for dwelling- and market-specific characteristics. More specifically, the multivariate Cox hazard model expresses the elasticities as "odds", e.g. a coefficient of 1.2 means a 1.2 times quicker "dead" as the reference or baseline. In a first step, we parametrize the equation in a semiparametric proportional hazard model:

$$\mathrm{h}(\mathrm{t}_{\mathrm{ijp}}) = \exp(\mathbf{X}_{\mathrm{ip}}\beta + \mathbf{Z}_{\mathrm{j}}\alpha + \boldsymbol{\mu}_{\mathrm{ip}}\delta_{\mathrm{p}} + \boldsymbol{\mu}_{\mathrm{ij}}\rho_{\mathrm{j}}) + \mathrm{e}_{\mathrm{ijp}}\ \forall\ \mathrm{m};\ \mathrm{m} \in 1, \dots, 7 \tag{3}$$

where h corresponds to hazard function of time-on-market t, the $\mathbf{X}$ matrix contains the specific characteristics of dwelling i at observation period p, $\mathbf{Z}$ includes socioeconomic data on ZIP-area j and $\boldsymbol{\mu}_{\mathrm{ip}}$ and $\boldsymbol{\mu}_{\mathrm{ij}}$ account for p time- and j spatial effects respectively. The results of equation (3) are expected to provide information on the covariates boosting or limiting the marketing time of dwellings in the observed housing markets.

A second step captures the spatial effects liquidity. While the covariates in $\mathbf{X}$ and $\mathbf{Z}$ are either continuous or binary, the p time- and j spatial effects in the matrices $\boldsymbol{\mu}_{\mathrm{ip}}$ and $\boldsymbol{\mu}_{\mathrm{ij}}$ are defined as follows:

$$\mu_{\mathrm{ip}} = \{1 \Leftrightarrow \mathrm{i\ in\ p}; 0 \Leftrightarrow \mathrm{else}\} \tag{4}$$

$$\mu_{\mathrm{ij}} = \{1 \Leftrightarrow \mathrm{i\ in\ j}; 0 \Leftrightarrow \mathrm{else}\} \tag{5}$$

For each m, the vector of $\hat{\rho}_{\mathrm{j}}$ coefficients captures the ZIP-specific relative changes in liquidity over the entire observation period with respect to certain reference category. The reference category in each market is defined as the ZIP-area with the highest asking rent R adjusted for sample size. Afterwards the results of the $\hat{\rho}_{\mathrm{j}}$ coefficients are presented in maps to explore liquidity graphically in a spatial context.

In a third step, equation (3) is expanded by non-linear effects. This improves the estimates of two continuous hedonic covariates: dwellings' rent and age. This is accomplished by applying a non-parametric smoothing estimator, which corresponds to a penalized approach comprehending of k knots. In simple words, for $k = 2$ the smoothing estimator minimizes the sum of squares of a "line" with one turning point or local minima, similar to quadratic terms. The knots are chosen iteratively by minimizing the sum of squares at different values of k (Heckman and Ramsay, 2000). The expanded Cox hazard model is as follows:

$$\mathrm{h}(\mathrm{t}_{\mathrm{ijp}}) = \exp\left(\mathbf{X}_{\mathrm{ip}}\beta + \mathbf{Z}_{\mathrm{j}}\alpha + \boldsymbol{\mu}_{\mathrm{ip}}\delta_{\mathrm{p}} + \boldsymbol{\mu}_{\mathrm{ij}}\rho_{\mathrm{j}} + f(\mathrm{x}_{\mathrm{ip}}^{\mathrm{a}}) + f(\mathrm{x}_{\mathrm{ip}}^{\mathrm{b}})\right) + \mathrm{e}_{\mathrm{ijp}}\ \forall\ \mathrm{m};\ \mathrm{m} \in 1, \dots, 7 \qquad (6)$$

,where $f(\mathrm{x}_{\mathrm{ip}}^{\mathrm{a}})$ and $f(\mathrm{x}_{\mathrm{ip}}^{\mathrm{b}})$ correspond to the smoothing function of dwellings' rent and age respectively. The coefficients are interpreted graphically, for each covariate in each market.

3. DATA DESCRIPTION AND STYLIZED FACTS

The estimation sample comprises two merged databases. On the one hand, 335'972 observations of rental flats are gathered from multiple listing services (MLS) in Germany from 2013-Q1 until 2016-Q3 as collected by the Empirica Systems database, which contains the most important multiple listing service (MLS) providers. On the other hand, two socioeconomic variables, the purchasing power per household and number of households on ZIP-code level, are extracted from the GfK-database. 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 calculated. Both variables might control for spatial distribution of dwellings within an urban area. NUTS3 regions correspond to the "Nomenclature of territorial units for statistics", which is a hierarchical system for dividing up the economic territory in Europe. While the NUTS1 consists on major socio-economic regions, the NUTS3 regions cover small regions similar to counties or administrative districts. (www.ec.europa.eu/eurostat/web/nuts/overview). Finally, the relevant variables in the context of hedonic survival regressions, dwellings' atypicality and the degree of overpricing are derived.

In Exhibit 1, Munich displays the highest density of ZIP-code areas with one ZIP-code each 4.14 km². Although the density of postal areas for the Dusseldorf subsample is higher than for Berlin, dwellings are on average located closer to the ZIP-code area centroid. The highest construction activity seems to have taken place in Munich, as the dwellings are on average only 36.16 years old and 20.9% are listed for first occupancy. A very low degree of atypicality – also lowest standard deviation in atypicality – together with the highest ratios for the amenities parking slot and elevator are likely signs for a large share of professional housing construction, meeting the demand of the wealthiest

households among the sample. Households in Munich have on average 57% higher purchasing power than households in Berlin, but pay 73.1% higher rent. Stuttgart and Munich have the most liquid markets for residential leasehold property as in both cities a dwelling is advertised about 4.3 weeks on average, while in Stuttgart the duration is slightly shorter and displaying lower standard deviation

------ Exhibit 1 ------

In order to conceptualize the main idea of survival methods and liquidity per se, a survival regression with city dummies is presented. Exhibit 2 shows the survival function proceedings from regressing time-on-market in weeks on city dummies over the entire sample period. The survival functions illustrate the mortality rate of an average dwelling as a function of time. When looking at the top 7 markets combined, the results show that the probability of letting a dwelling after one year is roughly 60%. Whereas the probability of letting an average dwelling in Munich and Stuttgart after one year is at about 70%, the probability of finding a new tenant in Dusseldorf within the first year is at only 50%. The lower panel shows, that the sharpest increase in the probability of letting an apartment in Munich, Stuttgart and Hamburg compared to the top 7 markets happens during the first three months, revealing the huge demand pressure which is resulting in above average liquidity for rental units within those cities. The inverse pattern appears for Dusseldorf and Frankfurt. Landlords in Cologne face lower than average liquidity within the first year, before the probability of letting a dwelling increases above the market average. As the survival functions evidently show distinctions between the cities, this paper aims at exploring the factors that boost or dampen the survival function in a multivariate approach.

------ Exhibit 2 ------

4. ECONOMETRIC RESULTS

4.1. Main liquidity drivers

Exhibit 3 presents the results of the three parameterizations from equation 3 for each city. The coefficients of the cox proportional hazards model are displayed together with their respective standard deviations, whereas positive coefficients increase the hazard (shorten the survival time) and increase therefore liquidity (dwelling's letting process). Since hazard models estimate event probabilities per units of time, a coefficient of determination just as in the OLS is difficult to obtain. As a substitute, the Pseudo-R^2 based on Kendall's Tau measures the concordance between estimated survival time and the observed survival time for only the non-censored response sample. Values between 80 % and 60 % are common in survival studies. Model I includes only hedonic covariates, whereas model II includes the

atypicality and overpricing indices as well as gravity variables and Cartesian coordinates. Model III presents the full model including all control variables.

The full models show, that liquidity responds negatively to rents and size. Thus, a dwelling's letting process is longer, the higher is the asking rent and the larger the dwelling, whereas the rent effect in Hamburg is insignificant. The factor age shortens the letting process in Berlin and Frankfurt, as the coefficients are significant and positive. In contrast, the design of the flats in terms of the number of rooms shows the expected effect as the higher the dwellings' usability, the shorter is the average letting process.

When focussing on the hedonic dummy variables, the coefficient interpretation is more tangible[i]. Dwellings with a bathtub, a parking slot, a balcony and an elevator are in general difficult to let, as the coefficients are in most of the cases negative. In Hamburg, the city with the highest ratio of flats with bathtubs, the feature seems to be accepted standard as it is the sole city revealing no significant impact. A similar effect is observed for built-in kitchens in Stuttgart. In Dusseldorf and Cologne on the other hand, where the least dwellings have built-in kitchens, the presence of a kitchen has a rather strong positive impact on liquidity. A different picture emerges for Munich, where across the different cities on average the most dwellings include a parking slot. Despite the high ratio of 65%, the feature cannot be declared as standard, as it shows a significant negative impact on liquidity. For terraces in Stuttgart and elevators in Munich it can be said that supply meets demand. The highest ratio for terraces is found in Stuttgart while across the cities the most dwellings with elevators are found in Munich. For both cities, the presence of these features decreases the marketing time. In each city besides Frankfurt, where the highest ratio of dwellings offered for first occupancy are located, the feature increases the liquidity.

The degree of atypicality and the degree of overpricing show consistent results, as they confirm a restricted liquidity for highly atypical and overpriced dwellings with exception of Munich and Stuttgart. Only there, the degree of overpricing has a positive impact on liquidity, probably attributable to the strong demand of the last years accompanied by insufficient housing construction, forcing households to let properties irrespective of the hedonic particularities. In addition to the strong overall demand for living space in Munich, there exists a forceful demand for dwellings in the heart of the city centre as the next chapter will demonstrate.

The spatial gravity variables included into the model show, that for dwellings in Frankfurt, Cologne and Hamburg the marketing time decreases with distance to the ZIP-code centroid. When looking at the spatial influence with larger granularity, it becomes clear that in six out of the seven largest German cities the proximity to the city centre is of

significant importance when marketing a dwelling. The coefficients display prolonged marketing time for more decentralised flats.

The different model parameterizations show a substantial change in the hazard rates for the asking rents vector after controlling for atypicality, overpricing and gravity variables and especially when including time, spatial and socioeconomic variables. The increase in the Pseudo R^2 between model I and II confirms however that liquidity is more accurately captured when controlling for the latter variables leading to a less pronounced bias from omitted variables.

------ Exhibit 3 ------

4.2. 2 Liquidity in an urban spatial context

When looking at the distribution of Berlin's most liquid ZIP-regions, one can clearly detect a more or less circular form surrounding the inner city. As it is impossible to infer a pattern of spatial preferences, the cause behind the strong demand for more decentralised dwellings might simply be the rental aspect. As central Berlin is getting more expensive, lower income households have to resettle to the more affordable outskirts. A very distinct constellation appears when looking at the most liquid regions of Hamburg. Since almost the whole southern half of Hamburg is enclosed by widespread natural reserves and Europe's third largest harbour, only solitary settlements are found within that area. The ZIP-regions with very high liquidity cluster themselves west- and eastwards of the old town. Especially some of the western districts are among the most densely populated areas of Germany. Cologne shows a constant liquidity pattern for almost all of its ZIP-code regions. In Munich especially the more expensive inner-city regions are of very high liquidity and let faster than the reference district. Those central districts contain on average 14% less parking slots than the city average, thus explaining the negative effect on liquidity when looking at the whole city. The expansion of Stuttgart's inner-city is naturally bound by it's geology. The districts displaying the highest market liquidity are found south of the kettle containing the inner-city.

------ Exhibit 4a ------

------ Exhibit 4b ------

4.3. 3 Non-linearity and its impact on liquidity

In this section, the effects of non-linear covariates on liquidity are presented. More specifically, the basic equation (3) is expanded by smoothing functions of rent and age as described in equation (6). Based on the results from exhibit 3,

the non-linear effects of rent and age are presented in exhibit 5 as log hazards. In this context, values above zero increase the hazard, i.e. shorten the survival time, and increase therefore liquidity, i.e. dwelling's letting process[ii].

------ Exhibit 5a ------

------ Exhibit 5b ------

The graphs evidently show that liquidity responds non-linear to the dwelling asking rent and age. Besides for Munich where less expensive dwellings are always let faster, the log hazard functions display a kinked shape showing a short increase in liquidity with rent before the market liquidity declines with higher rent. Each city exhibits its individual threshold at which liquidity is maximised. The non-linear results of age show a pronounced liquidity discount for dwellings smaller than ca. 60 m² in Berlin and 40 m² in Frankfurt. In Stuttgart and Cologne on the other hand liquidity discounts for flats larger than 80m² and 50m² respectively.

5. Conclusions

A common understanding of liquidity and the underlying factors is essential when assessing market movement or selling and buying property by market players. This paper explored several concepts of liquidity in residential rental markets and introduced a profound base on the econometric tools necessary to capture liquidity. The results based on big data assist both private and institutional landlords in assessing the marketability of rental property within the observed markets and help governmental organisations in charge of housing and urban planning when deriving political implications. Across the seven largest German real estate markets the semiparametric cox-hazard models controlling for hedonic, socioeconomic, spatial and various fixed effects displayed similarities as well as differences in the liquidity and its determinants. While for each city the asking rent, living area, dwelling's age and distance to the NUTS3 centroid show consistent effects, the hedonic characteristics and the degree of overpricing display market specific impact on liquidity. Based on those results geographic liquidity patterns are derived for the observed cities and individual non-linear effects of asking rent and dwelling's age are shown graphically. The paper contributes to a better understanding of liquidity in rental markets, a less explored topic in traditional (housing) real estate research.

6. Literature

Allen, M.T., Rutherford, R.C. and Thomson, T.A. (2009), "Residential Asking Rents and Time on the Market", *Journal of Real Estate Finance and Economics*, Vol. 38 No. 4, pp. 351-365.

Anglin, P.M., Rutherford, R.C. and Springer, T.M. (2003), "The Trade-off Between Selling Price of Residential Properties and Time-on-the-Market: The Impact of Price Setting", *Journal of Real Estate Finance and Economics*, Vol. 26 No. 1, pp. 95-111.

Bourassa, S., Cantoni, E. and Hoesli, M. (2010), "Predicting House Prices with Spatial Dependence: A Comparison of Alternative Methods", *Journal of Real Estate Research*, Vol. 32 No. 2, pp. 139-159.

Bourassa, S., Haurin, D.R., Haurin, J.L., Hoesli, M. and Sun, J. (2009), "House Price Changes and Idiosyncratic Risk: The Impact of Property Characteristics", *Real Estate Economics*, Vol. 37 No. 2, pp. 259-278.

Cajias, M., Fuerst, F. and Bienert, S. (2016), "Are Energy Efficiency Ratings Ignored in the German Housing Market? – Evidence from a Large-Sample Hedonic Study", Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2799206 (accessed 22 December 2016).

Cirman, A., Pahor, M. and Verbic, M. (2015), "Determinants of Time on the Market in a Thin Real Estate Market", *Inzinerine Ekonomika-Engineering Economics*, Vol. 26 No. 1, pp. 4-11.

Cox, D.R. (1972), "Regression Models and Life-tables", *Journal of the Royal Statistics Society*, Series B Vol. 34 No. 2, pp. 187-220.

Fik, T.J., Ling, D.C. and Mulligan, G.F. (2003), "Modeling Spatial Variation in Housing Prices: A Variable Interaction Approach", *Real Estate Economics*, Vol. 31 No. 4, pp. 623-646.

Glower, M., Haurin, D. and Hendershott, P.H. (1998), "Selling Time and Selling Price: The Influence of Seller Motivation", *Real Estate Economics*, Vol. 26 No. 4, pp. 719-740.

Goodman, A.C. and Thibodeau, T.G. (2007), "The Spatial Proximity of Metropolitan Area Housing Submarkets", *Real Estate Economics*, Vol. 35 No. 2, pp. 209-232.

Haurin, D.R. (1988), "The Duration of Marketing Time of Residential Housing.", Real Estate *Economics*, Vol. 16 No. 4, pp. 396-410.

Haurin, D.R., Haurin, J.L., Nadauld, T. and Sanders, A. (2010), "List Prices, Sale Prices and Marketing Time: An Application to U.S. Housing Markets", *Real Estate Economics*, Vol. 38 No. 4, pp. 659-685.

Haurin, D.R., McGreal, S., Adair, A., Brown, L. and Webb, J.R. (2013), "List price and sales prices of residential properties during booms and busts", *Journal of Housing Economics*, Vol. 22 No. 1, pp. 1-10.

Heckman, N.E. and Ramsay, J. (2000), "Penalized regression with model-based penalties", *Canadian Journal of Statistics*, Vol. 28 No. 2, pp. 241-258.

Hoeberichts, M., van Rooij, M. and Siegmann, A. (2013), "House List Prices and Durations in Boom and Bust" available at: https://ssrn.com/abstract=2321521 (accessed 16 Dezember 2016).

Kluger, B.D. and Miller, N.G. (1990), "Measuring Real Estate Liquidity", *Real Estate Economics*, Vol. 18 No. 2, pp. 145-159.

Krainer, J. (1999), "Real Estate Liquidity", *Economic Review*, Vol. 3 No. 1, pp. 14-26.

Pavlov, A.D. (2000), "Space-Varying Regression Coefficients: A Semi-parametric Approach Applied to Real Estate Markets", *Real Estate Economics*, Vol. 28 No. 2, pp. 249-283.

Rutherford, R.C., Springer, T.M. and Yavas, A. (2005), "Conflicts between principals and agents: evidence from residential brokerage", *Journal of Financial Economics*, Vol. 76 No. 3, pp. 627-665.

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.

Turnbull, G.K. and Dombrow, J. (2006), "Spatial Competition and Shopping Externalities: Evidence from the Housing Market", *Journal of Real Estate Finance and Economics*, Vol. 32 No. 4, pp. 391-408.

Voigtländer, M. (2009), "Why is the German Homeownership Rate so Low?", *Housing Studies*, Vol. 24 No. 3, pp. 355-372.

7. Notes

i) The interpretation of continuous variables is less sizeable as hazards do not allow a direct interpretation as in OLS models

ii) The interpretation varies only with respect to the reference as the log and antilog are proportionally related. The Package Survfit in R allows only the visualization of non-linear effects in the base model.

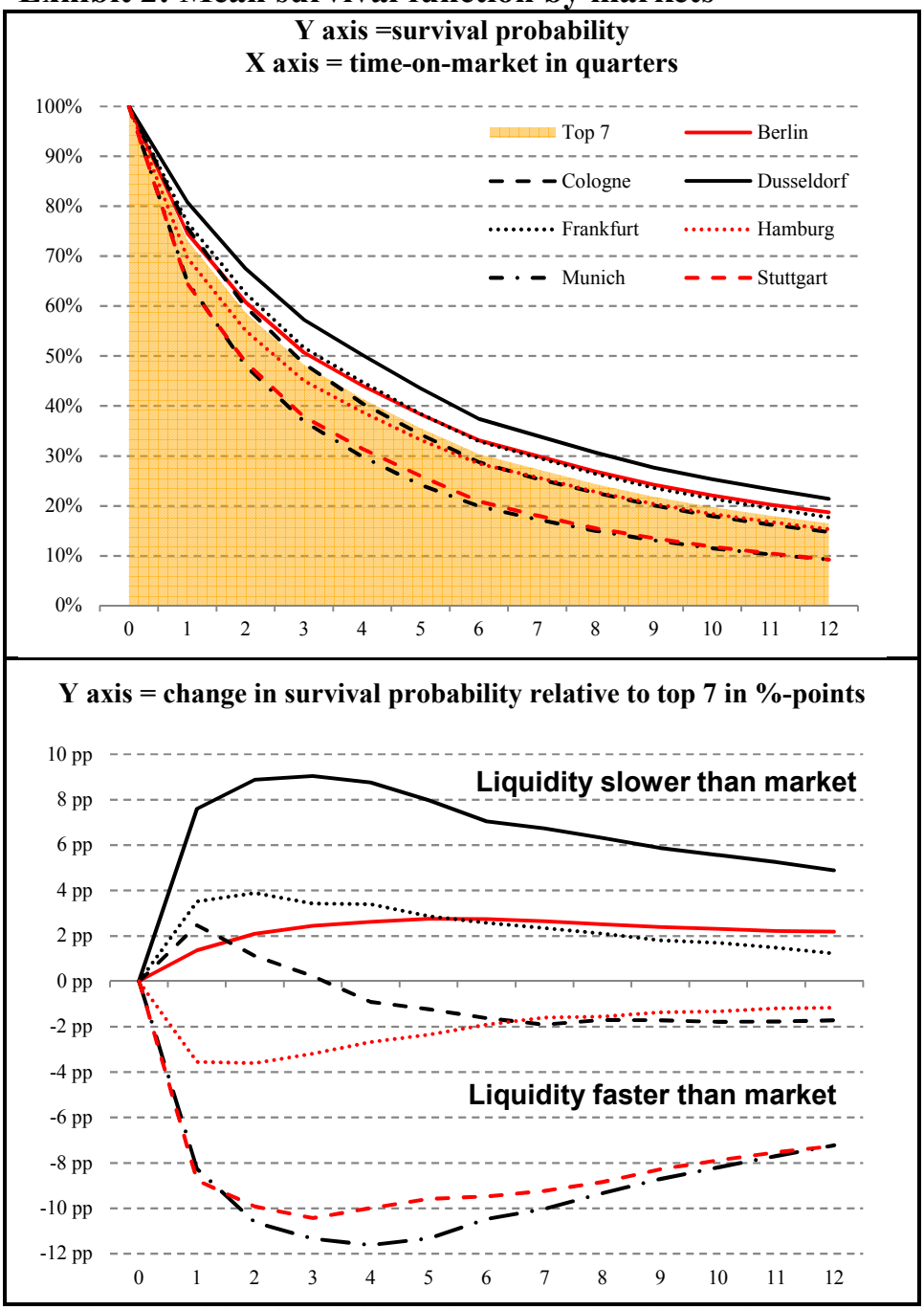
8. Exhibits

Exhibit 1: Descriptive Statistics

Variable Mean / (St. Deviation)	Berlin	Frankfurt	Munich	Stuttgart	Cologne	Dusseldorf	Hamburg
Asking rent €/m²/p.m.	8.567 (2.481)	12.226 (2.873)	14.831 (3.252)	11.06 (2.505)	9.788 (2.147)	9.599 (2.125)	10.567 (2.754)
Time-on-market weeks	6.71 (10.899)	6.564 (10.292)	4.331 (8.751)	4.316 (7.678)	5.996 (10.368)	7.717 (11.888)	5.734 (9.721)
Area m²	72.192 (27.939)	76.179 (31.489)	74.774 (31.475)	78.473 (31.244)	71.684 (27.746)	74.453 (29.78)	70.545 (26.988)
Age	56.139 (35.432)	40.549 (33.277)	36.159 (27.586)	44.776 (31.448)	41.803 (26.501)	49.55 (27.811)	46.308 (30.787)
Euclidean distance to ZIP centroid in Km.	0.687 (0.416)	0.748 (0.492)	0.601 (0.374)	0.717 (0.469)	0.915 (0.579)	0.652 (0.483)	0.808 (0.577)
Euclidean distance to NUTS3 centroid in Km.	8.466 (4.035)	3.677 (1.942)	4.733 (2.069)	4.437 (2.33)	5.462 (2.841)	3.958 (2.609)	7.974 (3.772)
Degree of atypicality	3.906 (3.251)	0.219 (0.165)	0.106 (0.1)	1.063 (0.375)	0.375 (0.106)	-0.101 (0.153)	0.12 (0.285)
Degree of overpricing	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)	0 (1)
Number of households in ZIP	11’856 (3564.845)	10’963 (4201.319)	11’402 (3088.482)	10’271 (3156.56)	13’066 (3393.028)	9’599 (2946.49)	10’960 (3360.45)
Purchasing power per household in €	34’791 (4’924)	46’941 (6’320)	54’618 (6’072)	46’901 (4’428)	45’589 (5’776)	47’658 (5’881)	42’804 (7’644)
With bathtub	0.607 (0.488)	0.579 (0.494)	0.538 (0.499)	0.49 (0.5)	0.482 (0.5)	0.531 (0.499)	0.627 (0.484)
With built-in-kitchen	0.474 (0.499)	0.662 (0.473)	0.667 (0.471)	0.685 (0.464)	0.297 (0.457)	0.359 (0.48)	0.767 (0.423)
With parking slot	0.258 (0.438)	0.524 (0.499)	0.65 (0.477)	0.575 (0.494)	0.424 (0.494)	0.394 (0.489)	0.408 (0.491)
With terrace	0.156 (0.363)	0.152 (0.359)	0.19 (0.392)	0.21 (0.408)	0.171 (0.376)	0.163 (0.369)	0.165 (0.371)
With balcony	0.668 (0.471)	0.642 (0.479)	0.638 (0.481)	0.616 (0.486)	0.627 (0.484)	0.637 (0.481)	0.659 (0.474)
With elevator	0.425 (0.494)	0.458 (0.498)	0.55 (0.497)	0.314 (0.464)	0.355 (0.479)	0.36 (0.48)	0.286 (0.452)
First occupancy?	0.179 (0.384)	0.24 (0.427)	0.209 (0.406)	0.188 (0.391)	0.152 (0.359)	0.168 (0.374)	0.164 (0.37)
N	119’481	28’641	32’216	12’755	36’940	35’814	70’125

Notes: The sample contains 335,972 observations of dwellings advertised on multiple listing services (MLS). The sample covers 3.75 years from Q1 2013 until Q3 2016. While the asking rent is expressed in €/m²/month, area is expressed as m² and age as a number of years, the means of the characteristics bathtub, built-in kitchen etc. can be interpreted as ratios. The purchasing power per household and the number of households per ZIP-code area are extracted from the market research database of GfK. Spatial gravity indicators measure the Euclidian distance of each dwelling to the geographical centroid of the ZIP and NUTS3 polygon in kilometers. The degree of atypicality is calculated according to the definition by (Haurin, 1988), while the Degree of overpricing is constructed according to (Anglin et al., 2003).

Exhibit 2: Mean survival function by markets



Notes: The upper panel shows the survival function from a cox regression of dwellings' time-on-market in weeks on seven dummies (stratas) and the entire sample. The survival functions illustrate the mortality rate of an average dwelling as a function of time. The lower exhibit presents the survival functions relative to the overall market survival.

Exhibit 3: Semiparametric proportional cox-hazard regression – Y=Time-on-market in weeks

Exp(coefficient) **Mean / (St. Deviation)**	Berlin			Frankfurt			Munich			Stuttgart			Cologne			Dusseldorf			Hamburg		
Model	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III
Log rent	-0,056 0,002***	0,016 0,003***	-0,045 0,007***	-0,053 0,001***	-0,014 0,003***	-0,035 0,006***	-0,048 0,002***	0,002 0,005	-0,11 0,009***	-0,054 0,005***	0,088 0,012***	-0,093 0,017***	-0,019 0,003***	0,09 0,007***	-0,085 0,014***	-0,052 0,003***	0,051 0,007***	-0,035 0,011***	-0,045 0,002***	0,036 0,005***	0,003 0,008
Log area	-0,008 0***	-0,01 0***	-0,014 0***	-0,011 0***	-0,012 0***	-0,015 0,001***	-0,011 0***	-0,011 0***	-0,017 0,001***	-0,01 0,001***	-0,01 0,001***	-0,016 0,001***	-0,009 0***	-0,009 0***	-0,015 0,001***	-0,008 0***	-0,01 0***	-0,011 0***	-0,009 0***	-0,011 0***	-0,014 0***
Age	0,002 0***	0 0***	0,012 0,001***	0,001 0***	0,001 0***	0,014 0,003***	0,001 0***	0,001 0***	0,003 0,002	0,001 0	0,001 0	-0,005 0,004	0,001 0***	0 0	-0,005 0,002**	0 0	0 0	0 0,002	0,001 0***	0 0	0,002 0,002
Number of rooms	0,123 0,006***	0,171 0,006***	0,199 0,007***	0,159 0,012***	0,174 0,013***	0,185 0,013***	0,144 0,012***	0,14 0,012***	0,203 0,015***	0,098 0,016***	0,099 0,017***	0,172 0,018***	0,116 0,011***	0,131 0,012***	0,175 0,013***	0,119 0,011***	0,14 0,012***	0,16 0,012***	0,126 0,009***	0,163 0,009***	0,172 0,01***
With bathtub (yes=1)	-0,159 0,007***	-0,177 0,007***	-0,167 0,007***	-0,11 0,014***	-0,094 0,014***	-0,082 0,014***	-0,056 0,013***	-0,07 0,013***	-0,024 0,013*	-0,129 0,02***	-0,136 0,02***	-0,053 0,021**	-0,071 0,011***	-0,071 0,012***	-0,042 0,012***	-0,118 0,012***	-0,122 0,012***	-0,102 0,012***	-0,022 0,009**	-0,026 0,01***	-0,012 0,01
With built-in kitchen (yes=1)	0,042 0,007***	-0,082 0,008***	0,044 0,013***	0,099 0,016***	-0,001 0,018	0,043 0,024*	0,018 0,014	-0,037 0,015**	0,071 0,019***	-0,011 0,022	-0,191 0,028***	0,062 0,043	0,096 0,013***	-0,022 0,014	0,162 0,019***	0,15 0,013***	0,037 0,015**	0,228 0,019***	-0,112 0,012***	-0,227 0,014***	-0,246 0,028***
With parking slot (yes=1)	-0,112 0,008***	-0,159 0,008***	-0,039 0,011***	-0,073 0,018***	-0,114 0,018***	-0,057 0,02***	-0,085 0,016***	-0,101 0,017***	-0,05 0,018***	-0,045 0,025*	-0,097 0,026***	0,013 0,028	-0,063 0,013***	-0,109 0,014***	-0,021 0,016	-0,04 0,015***	-0,066 0,015***	0,026 0,017	-0,174 0,01***	-0,197 0,011***	-0,061 0,013***
With terrace (yes=1)	-0,116 0,009***	-0,131 0,009***	0,011 0,012	-0,075 0,017***	-0,093 0,018***	-0,021 0,02	-0,113 0,016***	-0,147 0,016***	-0,011 0,021	-0,054 0,023**	-0,111 0,024***	0,088 0,032***	-0,074 0,015***	-0,144 0,015***	0,053 0,021**	-0,064 0,016***	-0,111 0,017***	0,109 0,024***	-0,049 0,011***	-0,042 0,012***	0,076 0,014***
With balcony (yes=1)	-0,053 0,007***	-0,042 0,007***	-0,03 0,008***	-0,034 0,015**	-0,051 0,015***	-0,04 0,016**	-0,06 0,014***	-0,061 0,014***	-0,022 0,014	-0,044 0,022**	-0,086 0,023***	-0,009 0,025	-0,035 0,012***	-0,044 0,013***	-0,012 0,014	-0,028 0,013**	-0,061 0,014***	-0,03 0,015**	0,023 0,01**	-0,003 0,01	0,021 0,011*
With elevator (yes=1)	0,019 0,007***	-0,05 0,008***	-0,063 0,009***	0,001 0,017	-0,059 0,018***	-0,035 0,019*	0,077 0,014***	0,011 0,015	0,035 0,016**	-0,055 0,027**	-0,129 0,028***	-0,066 0,03**	-0,083 0,013***	-0,132 0,014***	-0,151 0,014***	-0,046 0,013***	-0,117 0,014***	-0,04 0,016**	-0,064 0,011***	-0,094 0,012***	-0,162 0,013***
First occupancy (yes=1)	-0,208 0,009***	-0,36 0,011***	0,097 0,03***	-0,144 0,019***	-0,215 0,019***	0,047 0,036	-0,236 0,017***	-0,374 0,019***	0,096 0,044**	-0,117 0,029***	-0,343 0,034***	0,297 0,077***	-0,16 0,019***	-0,336 0,02***	0,139 0,041***	-0,113 0,018***	-0,245 0,02***	0,235 0,042***	-0,171 0,013***	-0,234 0,014***	0,155 0,025***
Degree of atypicality		-0,266 0,007***	-3,571 0,29***		-0,23 0,015***	-1,904 0,493***		-0,279 0,017***	-1,955 0,622***		-0,407 0,031***	-2,765 0,821***		-0,329 0,014***	-3,221 0,492***		-0,236 0,015***	-4,246 0,5***		-0,243 0,012***	-3,749 0,351***
Degree of overpricing		-0,042 0,008***	-0,104 0,019***		0,024 0,016	-0,058 0,032*		0,054 0,018***	0,177 0,036***		0,049 0,024**	0,164 0,05***		0,077 0,011***	-0,058 0,03*		-0,034 0,015**	-0,019 0,022		-0,019 0,009**	-0,002 0,025
Centroid to ZIP		-0,019 0,001***	-0,011 0,012		0,003 0,005	0,067 0,024***		-0,021 0,004***	-0,006 0,022		0,022 0,007***	-0,04 0,029		-0,003 0,003	0,065 0,015***		-0,014 0,003***	0,016 0,019		-0,015 0,002***	0,027 0,013**
Centroid to NUTS3			-0,056 0,007***			-0,07 0,018***			-0,046 0,016***			-0,076 0,021***			0,004 0,008			-0,13 0,015***			-0,086 0,007***
Coordinates		✓	✓		✓	✓		✓	✓		✓	✓		✓	✓		✓	✓		✓	✓
Socioeconomic variables			✓			✓			✓			✓			✓			✓			✓
Age fixed effects			✓			✓			✓			✓			✓			✓			✓
ZIP fixed effects			✓			✓			✓			✓			✓			✓			✓
Time fixed effects			✓			✓			✓			✓			✓			✓			✓
Smoothing functions																					
Pseudo R^2	61,9%	64,4%	65,9%	63,3%	64,3%	65,8%	65,3%	66,7%	68,1%	64,3%	66,2%	67,8%	60,9%	63,6%	65,3%	61,3%	62,6%	64,5%	64,9%	67,0%	68,2%
N	119’481			28’641			32’216			12’755			36’940			35’814			70’125		

Notes:*Significant at the 10%-level; ** significant at the 5%-level; *** significant at the 1%-level. The exhibit shows the regression results of a semiparametric cox regression of dwellings’ time-on-market in weeks on hedonic, spatial, socioeconomic and smoothing covariates. The results are presented as coefficients, while significant positive values shorten the survival and thus increase the assets’ liquidity, significant negative coefficients decrease the hazard rate and lengthen the survival. The three different model parameterizations control for different fixed effects. The Pseudo-R^2 based on Kendall’s Tau measures the concordance between estimated survival time and the observed survival time for only the non-censored response sample.

Exhibit 4a: Market liquidity conditions relative to ZIP with the highest adjusted rental level

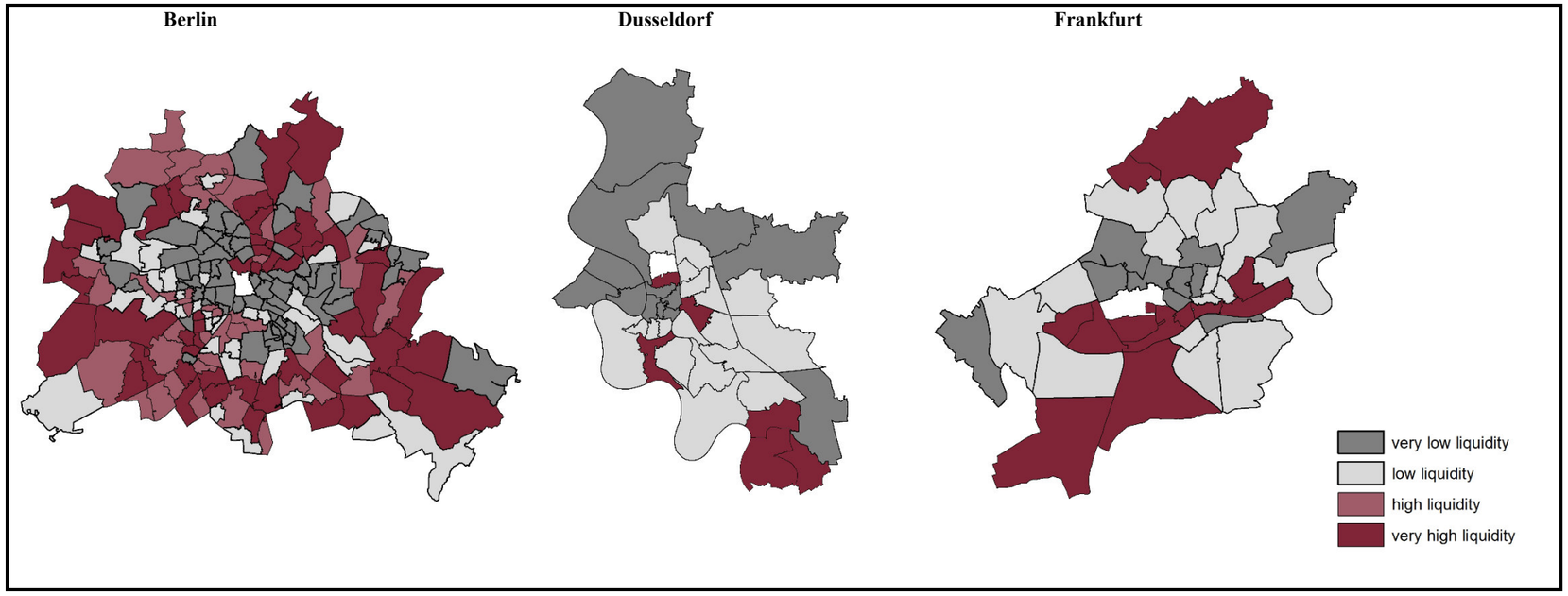
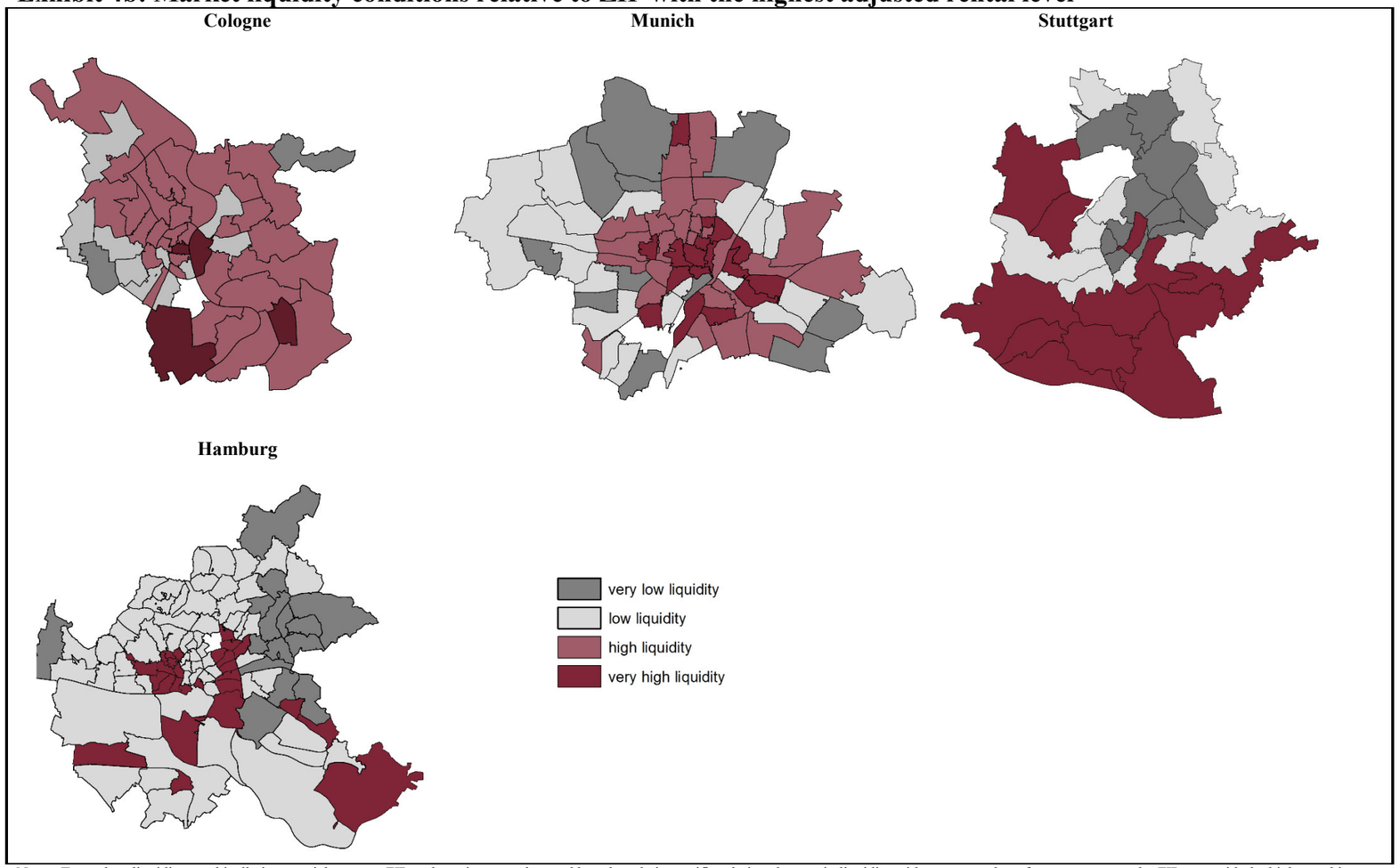


Exhibit 4b: Market liquidity conditions relative to ZIP with the highest adjusted rental level



Notes: To explore liquidity graphically in a spatial context, ZIP-code regions are clustered based on their specific relative changes in liquidity with respect to the reference category, the ZIP-area with the highest asking rent adjusted for sample size.

Exhibit 5a: Non-linear covariates' effect on dwellings' time-on-market hazard

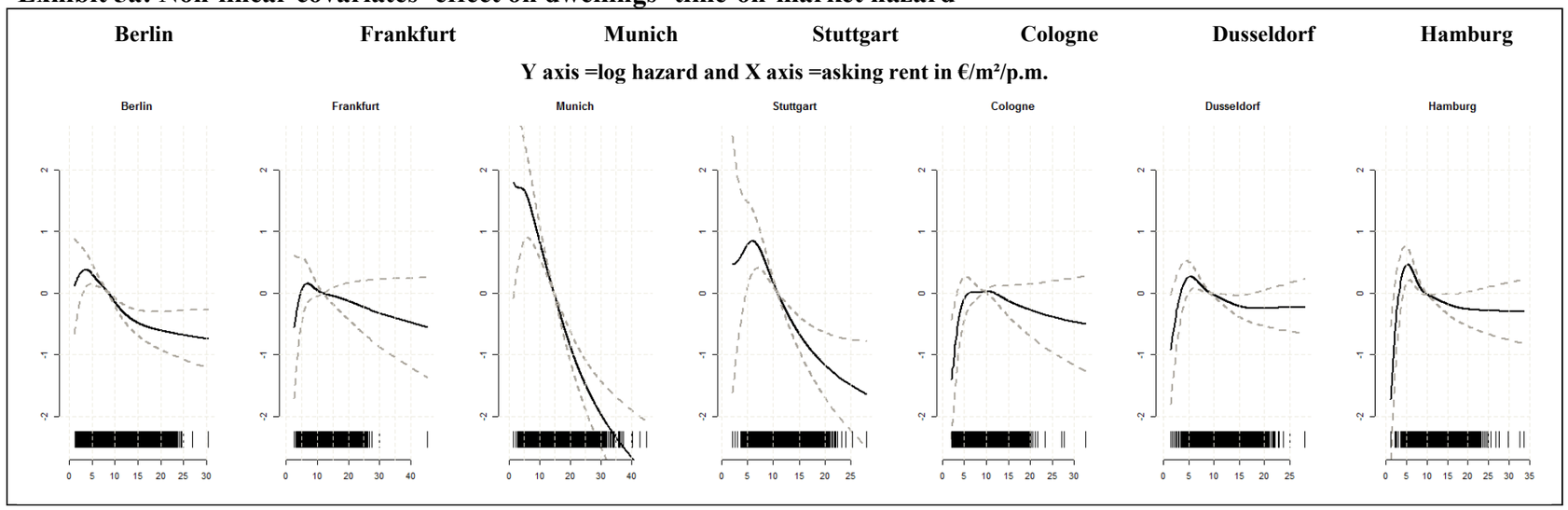
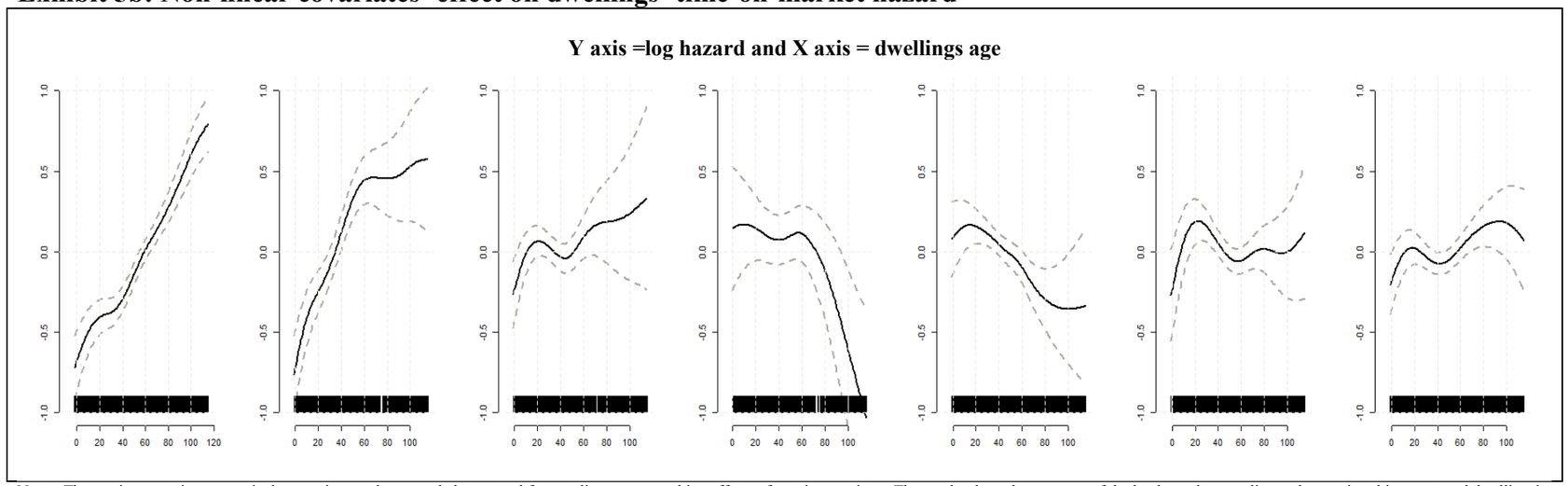


Exhibit 5b: Non-linear covariates' effect on dwellings' time-on-market hazard



Notes: The semiparametric cox survival regression can be expanded to control for non-linear or smoothing effects of metric covariates. The results show the response of the log hazard to non-linear changes in asking rents and dwellings' age. Values above zero increase the hazard and consequently asset's liquidity. The vertical bars above the x axis display the density of the rent-level and age respectively. Confidence intervals are shown dashed.