

# Statistical Modeling of Rent Per Square Meter in Munich City, Germany

Ugochukwu Onumadu 

Department of Educational Specialties (Socioscientific Studies), Austin Peay State University, Clarksville, TN, USA

Email: [1uonumadu@apsu.edu](mailto:1uonumadu@apsu.edu)

**How to cite this paper:** Onumadu, U. (2025) Statistical Modeling of Rent Per Square Meter in Munich City, Germany. *Journal of Applied Mathematics and Physics*, 13, 3016-3053. <https://doi.org/10.4236/jamp.2025.139172>

**Received:** August 8, 2025

**Accepted:** September 20, 2025

**Published:** September 23, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

This study explores a comprehensive statistical model for analyzing rental apartment prices per square meter in Munich, Germany. The research investigates key quantitative and qualitative variables influencing rent dynamics by leveraging a robust dataset comprising over 2.6 million apartments with 59 variables, sourced from FDZ Ruhr and ImmobilienScout24, for the years 2015 and 2019. Thirty-one key variables (9 quantitative and 22 qualitative) were analyzed, and the study identified significant predictors, such as apartment size, furnishing quality, energy efficiency, and amenity availability, through exploratory data analysis and multiple linear regression with nonlinear covariates. Applying log transformations and polynomial terms improved model performance, with the 2019 model achieving an adjusted R-squared of over 0.54 in the Analysis Of Variance (ANOVA) ratio tests. Model diagnostics, including the Akaike Information Criterion (AIC), residual plots, and Variance Inflation Factor (VIF), were employed to assess model fit and multicollinearity, ensuring the robustness and validity of the regression model. The results indicate a consistent trend where larger apartments and permitting pets command lower rent per square meter, while upscale furnishings, kitchens, and the number of bedrooms are associated with higher prices. This study provides meaningful predictive analytics insights into urban housing and Munich's evolving rental market. The findings provide valuable insights for real estate planning, sustainable housing policies, urban development strategies, and educators, particularly for university administrators and planners who can advocate for informed housing policies. This research contributes to academic literature on rent modeling and provides a data-driven foundation for evidence-based decision-making in high-demand urban housing markets.

## Keywords

Statistical Modeling, Munich Housing Market, Rent Price Modeling, Multiple Linear Regression

## 1. Introduction

### 1.1. Introduction with Munich Rental Apartment Review

The importance of using statistical methods to develop a mathematical equation that models the relationship between a response variable  $\text{rent\_sqm}$  and a set of explanatory variables can not be overemphasized. The demand for apartment rentals in Germany, especially in Munich and Berlin, is relatively high compared to other cities. Between 2011 and 2016, about 45,000 new apartments were built in Munich for roughly 90,000 people, even as the population in Munich rose from 200,000 to 1.55 million during the same period [1]. Therefore, about 55,000 more apartments were needed to accommodate the new arrivals. By 2030, about 150,000 apartments would be required as the population will increase to more than 1.7 million based on the estimate of [1]. Germany is representative of the situation in many respects compared to other high-income countries like the UK, France, the US, Canada, etc., and therefore, apartment prices and rents are causing serious problems as they have risen significantly in the country's large cities [2]. In international comparisons, like North America or Southern Europe, Germany has a higher share of renters. For instance, in 2018, the homeownership rate in Germany was 51.5% compared to 65.1%, 72.4%, and 96.4% in the UK, Italy, and Romania, respectively.

### 1.2. Objective

This paper models  $\text{rent\_sqm}$  in Munich using multiple linear regression to uncover key market trends. It also investigates whether a transformation of the response variable is needed, examines influential covariates, and identifies those significantly influencing rent prices. The findings are intended to inform housing policy and guide educational leadership in housing planning.

### 1.3. Literature Review

Regression models, often log-linear, help address skewness and variability in housing data and key predictors include furnishing quality, energy efficiency, and modernization status [3] [4]. Cross-national comparisons highlight differences between Germany's state-supported and the U.S.'s market-driven housing systems [5] [6]. Sustainability concerns, especially the impact of energy-efficient design, have also gained attention [7]. This study builds on existing literature by modeling Munich's rental market and linking statistical analysis to educational policy, with implications for improving student and faculty housing strategies [8] [9].

### 1.4. Research Questions

- **RQ1:** Is there any relationship between the response variable (rent per sqm) and the predictors?
- **RQ2:** Does the relationship between the response variable and the predictors

require a transformation to satisfy linear regression assumptions?

- **RQ3:** What are the key predictors (covariates) that significantly influence the rental price per square meter in Munich's housing market?

## 2. Methodology

### 2.1. Research Design

This study employs a quantitative research design to investigate the rental price per square meter in Munich's housing market using a multiple linear regression model with nonlinear covariates.

### 2.2. Data Collection

A secondary source of data collection was used for this study. The data was provided by the FDZ Ruhr at RWI (and ImmobilienScout24) institution. The ImmobilienScout24 GmbH, founded in 1998, deals with real estate properties in Germany. The data set contains 2,651,885 observations and 59 attributes from 2007 to 2020. The data description is done in Section 3.

### 2.3. Sample Selection and Data Filtering

We first selected the two cities (Munich and Berlin) that have the highest number of rental transactions. Thereafter, we chose the two years (2015 and 2019) based on the significant impact observed in the plotted scatter of years with rent prices. For instance, Munic 2015 was filtered using the R code (`dfm15 <-df %>% filter(city == "Munich", year == 2015)`). We conducted separate studies of the two cities in two different papers. We focus on the city of Munich (2015 and 2019) for this article and conducted separate studies of Berlin in another article. The number of rental properties contained in each data set (Munich 2015 and Munich 2019) is 14,449 and 17,776, respectively, as shown in **Table 3**.

### 2.4. Data Cleaning and Missing Values

During data cleaning, we changed the variable names from German to English and removed outliers using the Interquartile Range (IQR) method. The recorded missing values and the NAs were part of the labels for most categorical variables, as shown in **Table 2**.

### 2.5. Data Analysis: Multiple Linear Regression with Nonlinear Covariate

#### 2.5.1. Concept of a Multiple Linear Regression Model

Often, a relationship between two (or more) variables is found or suspected. Sometimes, one might be interested in investigating whether there is a relationship or trend between two or more variables, and if they are, how they are related. In regression, we want to model the relationship between the variable of interest (dependent or response variable), and other given variables (covariates or independent variables); see [10]. For instance, we may want to know whether a relationship

exists between the number of hours students read in a day (independent variable or covariate) and their performance in the examination (dependent or response variable). The goal of regression analysis is to determine the parameters of the linear function that best describes the joint distribution of the response variable and the covariates [11]. We note that the relationship among variables may be linear, nonlinear (quadratic, cubic, etc.), or non-existent at all, and may involve several independent variables. Thus, we need tools for an exploratory data analysis (EDA), which enables us to suggest useful model formulations before fitting specific regression models. We refer to multiple linear regression when several independent variables are involved and the response variable is continuous. In this study, we want to investigate the relationship between the rent per square meter in Munich charged for an apartment characterized by continuous and discrete covariates.

### 2.5.2. Model Formulation

In a regression analysis with a continuous response variable  $Y_i$  and  $p$  covariates or predictors  $X_{i1}, X_{i2}, \dots, X_{ik}$  which may be continuous or qualitative (ordinal or nominal) with  $n$  observations, let  $(y_i, \mathbf{x}_i^\top) := (y_i, x_{i1}, \dots, x_{ik})^\top$ ,  $i = 1, \dots, n$ ,  $k = p - 1$ , be a pair of the  $i$ th observation  $(y_i, \mathbf{x}_i^\top)$  of the random vector  $(Y_i, \mathbf{x}_i^\top)$ , where  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})^\top$ , then our objective is to analyze the effects of the covariates on the mean value of the response variable ( $\mu_i \equiv E[Y_i]$ ). The linear model models the response as a linear function of the predictors together with an error term, *i.e.*

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i \quad (2.1)$$

with mean  $E[Y_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$ .

**Definition 2.1.** The multiple linear regression model is defined as

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \epsilon_i, \quad i = 1, \dots, n, \quad (2.2)$$

where  $\epsilon_i$  is the random error variable,  $\beta_0$  is the intercept, and the  $k$  parameters  $\beta_1, \dots, \beta_k$  are the unknown regression parameters to be estimated from  $n$  observations  $(y_i, x_{i1}, \dots, x_{ik})$ , for  $i = 1, \dots, n$ .

### 2.5.3. Polynomial Regression

Polynomial regression is often appropriate when a relationship exists between the response and the covariates. Given a continuous covariate  $V_i$  with observations  $v_i$  that has a polynomial effect of degree  $d$  on the response, then the model  $Y_i = \beta_0 + \beta_1 V_i + \beta_2 V_i^2 + \dots + \beta_d V_i^d + \dots + \epsilon_i$  can be used. Note, it is a linear regression model of the form (2.2) with  $x_{ij} = v_i^j$ ,  $j = 1, \dots, d$  [12] and [13].

In order to increase numerical stability, we orthonormalize the corresponding

design matrix  $X = \begin{pmatrix} 1 & v_1 & v_1^d \\ \vdots & \vdots & \vdots \\ 1 & v_n & v_n^d \end{pmatrix}$  to  $X^*$ , where all columns have unit norms and

are orthogonal. In  $R$ , this is achieved by  $\text{poly}(v, d)$ , see [14].

### 2.5.4. Transformations of the Response Variable

Sometimes, the transformation of the response variable is appropriate when non-normality and/or unequal error variances are present in the data. Let  $Y_i^{ln} := \ln(Y_i)$ , then the formulated model  $Y_i = \exp(\beta_0 + \beta_1 x_{i1}, \dots, \beta_k x_{ik} + \varepsilon_i)$  can be expressed in the form of the linear regression model (2.2) as

$$Y_i^{ln} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \dots, n \tag{2.3}$$

### 2.6. Estimation of Model Parameters

In this section, we will consider the methods of estimating the unknown parameters in the linear regression model of Definition (2.2). Our goal is to determine estimates

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \dots, \hat{\beta}_k)^\top \in \mathbb{R}^p \tag{2.4}$$

and the error variance  $\sigma$  based on  $n$  observations. Here  $\boldsymbol{\beta}$  is the unknown regression parameter vector.

Note that parameter **estimators**, which are random quantities are different from their realizations called **estimates**, which are determined by the values of the observations. We will consider two approaches: Least Squares (LS) estimation, and Maximum Likelihood (ML) estimation. These two estimation methods yield the same estimator if the assumptions of independence, homoscedasticity, and normality of errors are satisfied.

#### 2.6.1. Least Squares Estimation Method

Let the fitted values of the Model (2.2) be given as

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik}, \quad i = 1, \dots, n \\ &= \mathbf{x}_i' \hat{\boldsymbol{\beta}} \end{aligned} \tag{2.5}$$

Also, let the residual be denoted by  $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)^\top \in \mathbb{R}^n$ , which is the difference between the observed response values  $y_i$  and the corresponding fitted values of (2.11), be given as

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}, \tag{2.6}$$

where  $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_n)^\top \in \mathbb{R}^n$  in the vector notation. Then, least squares minimizes the residual sum of squares (the sum of the squared deviations) of Equation (2.12).

**Definition 2.2.** (Sum of squared deviations) Given the data  $(y_i, x_i), i = 1, 2, \dots, n$ , the sum of the squared deviations which is used in obtaining the estimates  $\hat{\boldsymbol{\beta}}$  of Equation (2.10) for the unknown regression parameters  $\boldsymbol{\beta}$  is given as

$$Q_{LS}(\boldsymbol{\beta}) = \sum_{i=1}^n (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 = \sum_{i=1}^n \hat{\varepsilon}_i^2 = \hat{\boldsymbol{\varepsilon}}^\top \hat{\boldsymbol{\varepsilon}} \tag{2.7}$$

In order to minimize  $Q_{LS}(\boldsymbol{\beta})$  (2.13), we take the partial derivative of  $Q_{LS}(\boldsymbol{\beta})$  with respect to  $\boldsymbol{\beta}$  and set the result to zero. Then, it follows

$$\frac{\partial(Q_{LS}(\boldsymbol{\beta}))}{\partial\boldsymbol{\beta}} = \mathbf{0} \Leftrightarrow -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \Leftrightarrow \mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y} \quad (2.8)$$

We are now interested in solving the least squares normal equations given in (2.14). If the matrix  $\mathbf{X}$  has a full rank  $p$ , then  $\mathbf{X}^T\mathbf{X}$  will be positive definite and will have a unique solution. Thus, the minimum of  $Q_{LS}(\boldsymbol{\beta})$  is attained at

$$\hat{\boldsymbol{\beta}}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \quad (2.9)$$

which is the least squares estimate from the normal equations.

### 2.6.2. Maximum Likelihood Estimation Method

The method of maximum likelihood estimation is based on specifying the distribution we are sampling from and writing the joint density of our sample, unlike in the least squares method where we do not specify the distribution of the response variable  $Y_i$ . Considering the assumptions of our linear model, we assumed in Equation (2.4) that the random variables  $Y_i$  are normally distributed ( $Y \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$ ). Thus, it follows that the likelihood of the vector  $(\boldsymbol{\beta}, \sigma)$  given the data values  $\mathbf{y}$  is

$$L(\boldsymbol{\beta}, \sigma | \mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \quad (2.10)$$

Therefore, the corresponding log likelihood is given by

$$l(\boldsymbol{\beta}, \sigma | \mathbf{y}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (2.11)$$

To maximize this log-likelihood (2.17) with respect to  $\boldsymbol{\beta}$ , we differentiate Equation (2.17) with respect to  $\boldsymbol{\beta}$  and set it equal to zero [15]. Thus, we have

$$\frac{\partial(l(\boldsymbol{\beta}, \sigma | \mathbf{y}))}{\partial\boldsymbol{\beta}} = \mathbf{0} \Leftrightarrow -\frac{1}{2\sigma^2}(-2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}) = \mathbf{0} \Leftrightarrow \mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y} \quad (2.12)$$

This shows that  $\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{LS}$ .

Also, differentiating Equation (2.17) with respect to  $\sigma^2$  and maximizing over  $\sigma^2$ , we have

$$\hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n} \|\hat{\boldsymbol{\epsilon}}\|^2 \quad (2.13)$$

and an unbiased estimator  $s^2$  of  $\sigma^2$  is given by

$$s^2 := \frac{1}{n-p} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \frac{n}{n-p} \hat{\sigma}^2 = \frac{1}{n-p} \|\hat{\boldsymbol{\epsilon}}\|^2. \quad (2.14)$$

### 2.6.3. Goodness of Fit and Model Selection

It is of great importance to know the goodness of the fitted model after estimating the parameters of the linear regression model of (2.2). Thus, we need suitable measures of the goodness of fit. Therefore, we will introduce one of the appropriate measures of the goodness of fit called the coefficient of determination ( $R^2$ ), which determines the proportion of variation of the response variable that is ex-

plained by the covariates.

### 2.6.4. Sum of Squares

**Definition 2.3. (Sum of squares)** We define the sum of squares SST (**total sum of squares**), SSR (**regression sum of squares**) and SSE (**error sum of squares**) to quantify the amount of variability explained by the regression model as follows

$$\begin{aligned} \text{SST} &:= \sum_{i=1}^n (y_i - \bar{y})^2 \Leftrightarrow (\text{total sum of squares}) \\ \text{SSR} &:= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \Leftrightarrow (\text{regression sum of squares}) \\ \text{SSE} &:= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \Leftrightarrow (\text{error sum of squares}) \end{aligned} \tag{2.15}$$

where  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ . Thus, we can have the decomposition as

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{2.16}$$

and using the fact that  $\sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$ , it follows from (2.26) that

$$\text{SST} = \text{SSR} + \text{SSE} \tag{2.17}$$

### 2.6.5. Selection of Model ( $R^2$ and Adjusted $R^2$ )

The multiple coefficient of determination  $R^2$  is a measure of goodness of fit. It measures how well the covariates in the model explain the variance in the response variable, see [16].

**Definition 2.4. (Multiple coefficient of determination)** We define the **multiple coefficient of determination  $R^2$**  as

$$R^2 := \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}} \tag{2.18}$$

We also define the **adjusted multiple coefficient of determination  $R_{adj}^2$**  as

$$R_{adj}^2 := 1 - \frac{\text{SSE}/(n-p)}{\text{SST}/(n-1)} \tag{2.19}$$

The values of the multiple coefficient of determination range from zero to one ( $0 \leq R^2 \leq 1$ ). Our model accounts for a larger variation of the response when the  $R^2$  is closer to 1. However, the weakness of  $R^2$  is that, it always increases when we add more covariates to our model, and therefore cannot be used to compare the goodness of fit for models with different numbers of covariates, see [17]. Thus, there is a need to establish an appropriate measure  $R_{adj}^2$  that compares models with different numbers of covariates. We will therefore make use of the adjusted multiple coefficient of determination ( $R_{adj}^2$ ) as a measure of our model selection in this paper.

### 2.6.6. Correlation Analysis

To measure the strength and direction of the linear relationship between two continuous variables, we use the correlation analysis. The most commonly used met-

ric is the Pearson correlation coefficient, denoted by  $\rho$  for the population and  $r$  for the sample. It ranges from  $-1$  to  $1$ , where values close to  $1$  or  $-1$  indicate strong positive or negative linear relationships, respectively, and values near  $0$  suggest no linear relationship.

The sample Pearson correlation coefficient between two variables  $X$  and  $Y$  is given by:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

where  $\bar{X}$  and  $\bar{Y}$  are the sample means of  $X$  and  $Y$ , respectively. This metric provides a preliminary indication of potential multicollinearity when applied to predictor variables.

## 2.7. Hypothesis Testing

A statistical hypothesis is an assumption about the form of a population, which based on sample information from the population, seeks to support or reject this assumption. If there is evidence that the null hypothesis (hypothesis of no difference) denoted by  $H_0$  is not true, then it is rejected and its alternative denoted by  $H_1$  is accepted. Thus, a test of hypothesis is a rule or a procedure used for deciding whether to accept or reject  $H_0$  or to determine whether the observed sample differs significantly from expected results under  $H_0$  [18]. This concept can be extended in statistical inference for the model parameters of linear regression [19]. For instance, we may want to know if the response variable is significantly influenced by a particular set of covariate variables, which can be expressed in terms of linear combinations of the unknown regression parameters  $\beta = (\beta_0, \dots, \beta_k)^\top$ . We will use the chi-square, F and the univariate t-distribution since the t-test and the F-test rely on quantities of these distributions.

**Definition 2.5. (Chi-square distribution)** A continuous random variable  $X$  is said to have a **Chi-square distribution** with parameter,  $\nu$ , if its probability density function is given by

$$f_X(x|\nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \nu > 0, x > 0$$

Here,  $\nu$  is the degree of freedom,  $E(X) = \nu$ ,  $\text{Var}(X) = 2\nu$ . Thus, we say that  $X$  follows a Chi-square distribution with  $\nu$  degree of freedom ( $X \sim \chi_\nu^2$ ).

**Definition 2.6. (F-distribution)** A continuous random variable  $X$  is said to have an **F-distribution** with degrees of freedom ( $df$ )  $\nu_1$  and  $\nu_2$ , if its pdf is given by

$$f(x) = \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right) \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1}}{\Gamma\left(\frac{\nu_1}{2}\right) \Gamma\left(\frac{\nu_2}{2}\right) \left(1 + \frac{\nu_1 x}{\nu_2}\right)^{\frac{\nu_1 + \nu_2}{2}}}, \quad x \geq 0. \quad (2.20)$$

If  $X_1 \sim \chi_{v_1}^2$  and  $X_2 \sim \chi_{v_2}^2$  and are independent, it follows in (2.30) that  $X$  is  $F$ -distributed with  $v_1$  and  $v_2$  df.

$$X = \frac{X_1/v_1}{X_2/v_2} \sim F_{v_1, v_2} \tag{2.21}$$

**Definition 2.7. (Univariate t-distribution)** A continuous random variable  $X$  is said to have a **Univariate t-distribution** with degree of freedom  $v$ , if its pdf is given by

$$f_v(x; \mu, \sigma^2) := \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v} \sigma} \left\{ 1 + \left(\frac{x-\mu}{\sigma}\right)^2 \frac{1}{v} \right\}^{-\frac{v+1}{2}}, \quad v \geq 1 \tag{2.22}$$

$$E(X) = \mu \text{ and } \text{Var}(X) = \frac{v}{v-2} \sigma^2.$$

If  $X_1 \sim N(0,1)$  and  $X_2 \sim \chi_n^2$  and are independent, it can be shown in (2.32) that  $T$  has a  $t$ -distribution with  $v$  df.

$$T = \frac{X_1}{\sqrt{\frac{X_2}{v}}} \sim t_v. \tag{2.23}$$

### 2.8. T-Test

**Definition 2.8. (t-test)** We define the **t-test procedure** for our model (2.2) as follows, since in a t-test, the test statistic is computed for each  $\beta_j$ , see [20].

**Hypotheses:**

$$H_0 : \beta_j = 0 \text{ versus } H_1 : \beta_j \neq 0$$

**Test statistic:**

$$T_j = \frac{\hat{\beta}_j}{\widehat{se}(\hat{\beta}_j)} \sim t_{n-p}, \text{ under } H_0 \tag{2.24}$$

Here,  $\widehat{se}(\hat{\beta}_j) := s \sqrt{\left[ (X^T X)^{-1} \right]_{jj}}$  is the estimated standard error of  $\hat{\beta}_j$  and

$s = \sqrt{s^2}$  defined in Equation (2.22)

**Rejection Rule:** Reject  $H_0$  at level  $\alpha$ , if  $|T_j| > t_{n-p, 1-\alpha/2}$

### 2.9. Analysis of Variance (ANOVA)

**Definition 2.9. ANOVA** is mostly used to summarize the hypothesis tests results in linear models in a tabular form. Given two models  $M_{\text{reduced}}$  and  $M_{\text{full}}$  which are nested:  $M_{\text{reduced}} \subset M_{\text{full}}$ , that is, all covariates of the reduced model are contained in the full model, we define the **ANOVA-test ratio** for the comparison of  $M_{\text{reduced}}$  and  $M_{\text{full}}$  as follows

$$F = \frac{(\text{SSE}_{\text{reduced}} - \text{SSE}_{\text{full}}) / (n - p_{\text{full}})}{\text{SSE}_{\text{reduced}} / (p_{\text{full}} - p_{\text{reduced}})} \sim F_{n-p_{\text{full}}, p_{\text{full}}-p_{\text{reduced}}} \tag{2.25}$$

## Hypotheses

$$H_0 : \beta_i = 0 \text{ versus } H_1 : \beta_i \neq 0$$

**Test statistic:**  $F$ , defined in Equation (2.25)

**Rejection Rule:** Reject  $H_0$  at level  $\alpha$ , if  $F > F_{(1-\alpha), n-p_{\text{full}}, p_{\text{full}}-p_{\text{reduced}}}$

## 2.10. Analysis of Residuals

After estimating the model parameters, the credibility of the assumptions of linearity, normality of errors, and homoscedasticity for the given data can be assessed using residuals. It is therefore important to study the residuals to examine the extent to which our model assumptions may be violated. Hence, investigating the patterns in the residual plots can help us determine if our model assumptions are violated or not. This is referred to as the analysis of residuals. Residual plots can help us decide whether to transform any of the covariates that we may want to include in the model or not.

## 2.11. Statistical Checks for the Plausibility of the Linear Model Assumptions

### 2.11.1. Linearity

The check we are going to use is the residuals versus the fitted values plot. If this plot has no trend, then we assume the linearity assumption as plausible [21].

### 2.11.2. Homoscedasticity

We are interested in checking if  $\text{Var}[Y_i] = \text{Var}[\varepsilon_i] = \sigma^2$  holds. To check this, we again used the standardized residual versus the residual plots. If the standardized residuals are not spread equally along the range of the fitted values, then we interpret the homoscedasticity assumption as not plausible, see [22].

### 2.11.3. Independence

To check if  $\text{Cov}(\varepsilon_j, \varepsilon'_j) = \rho = 0$  holds, we plot the residuals versus the covariates to see if the residuals are randomly and symmetrically distributed around zero. If this is true, we assume that the independence assumption is plausible [21].

### 2.11.4. Normality

To check for  $\varepsilon_i \sim N_n(0, \sigma^2 \mathbf{I}_n)$ , we use the Quantile versus Quantile plot (QQPlot). If we do not have a straight line on the QQ plots of our variable versus the theoretical normal quantile, then we assume that the normality assumption is not plausible [23].

### 2.11.5. Multicollinearity

To check for multicollinearity among explanatory variables  $X_1, X_2, \dots, X_p$ , we assess whether there is a strong linear relationship between them, which can inflate the standard errors of the estimated coefficients  $\hat{\beta}_j$ . This is commonly evaluated using the Variance Inflation Factor (VIF), defined as

$$\text{VIF}_j = \frac{1}{1 - R_j^2},$$

where  $R_j^2$  is the coefficient of determination from regressing  $X_j$  on the remaining predictors. A  $VIF_j > 5$  suggests a potentially problematic level of multicollinearity. Variables exceeding this threshold must be examined and removed if necessary to enhance model stability and interpretability [24] [25].

### 3. Data Description and Management

The data has both quantitative and qualitative covariates with rent per square meter (rent\_sqm) as the response variable. We focus on the most relevant 31 variables such as “the additional cost”, “heat cost”, “construction year”, etc. The quantitative covariates are summarized as follows: Min = Minimum, 25% = 1st quartile, 50% = Median,  $\bar{X}$  = Mean, 75% = 3rd quartile, Max = Maximum and Not available = NA. On the other hand, the qualitative covariates are summarized with their respective categories. Note that costs are expressed in EUR and rounded to two decimal digits and the following data summaries in **Table 1** and **Table 2** represent the whole data set.

**Table 1.** Description of quantitative variables.

Variables	Description
<b>rent_sqm</b>	Calculated rent per sqm by rent and size of apartment. Min = 3, 25% = 7, 50% = 9, $\bar{X}$ = 9.39, 75% = 12, Max = 28
<b>Addcost</b>	The extra monthly costs that need to be paid for other bills on top of the base rent excluding electricity. Min = 0, 25% = 100, 50% = 140, $\bar{X}$ = 153.8, 75% = 196, Max = 599, NA = 97,186
<b>Heatcost</b>	The monthly heating cost. Min = 0, 25% = 50, 50% = 70, $\bar{X}$ = 75.2, 75% = 94, Max = 300, NA = 898,984
<b>Conyear</b>	The year in which the object was built Min = 1851, 25% = 1930, 50% = 1970, $\bar{X}$ = 1964, 75% = 1996, Max = 2020, NA = 447,372
<b>Lmod</b>	The year of the last modernization Min = 1981, 25% = 2009, 50% = 2012, $\bar{X}$ = 2011, 75% = 2015, Max = 2018, NA = 1,113,056
<b>Lspace</b>	Living space in square meters Min = 19, 25% = 53, 50% = 68, $\bar{X}$ = 71.15, 75% = 85, Max = 165
<b>Fspace</b>	The usable floor space in square meters Min = 0, 25% = 16, 50% = 57, $\bar{X}$ = 54.8, 75% = 79, Max = 250, NA = 1,053,922
<b>Energycon</b>	The energy consumption per year and square meter in kWh Min = 0, 25% = 82, 50% = 117, $\bar{X}$ = 120.4, 75% = 152, Max = 350, NA = 977,343
<b>Adlength</b>	The difference between edat and adat. Min = 0, 25% = 0, 50% = 0, $\bar{X}$ = 0.71, 75% = 1, Max = 20

**Table 2.** Description of qualitative variables.

Variables	Description
<b>afloor</b>	Apartment-specific variable indicates the floor the apartment is located on. <b>afloorg</b> is used to group afloor as follows: (-1) - 0, 1 - 2, 3 - 9, >9, NA
<b>bfloor</b>	This indicates the number of floors in the building. <b>bfloorg</b> is used to group bfloor as follows: 0 - 2, 3, 4, 5, >5, NA
<b>nrooms</b>	Number of rooms, excluding kitchen, bath or corridors. <b>nroomsg</b> is used to group nrooms as follows: 1 - 1.5, 2 - 2.5, 3 - 3.5, >3.5, NA
<b>nbed</b>	Number of bedrooms of the property. <b>nbedg</b> is used to group nbed as follows: 0 - 1, 2, >2, NA
<b>nbath</b>	Number of bathrooms in the property <b>nbathg</b> is used to group nbath as follows: 0 - 1, >1, NA
<b>elevator</b>	This variable indicates if a property has an elevator. <b>elevatorg</b> is used to group elevators as follows: Yes, No, NA
<b>balcony</b>	This variable indicates the presence of a balcony. <b>balconyg</b> is used to group balcony as follows: Yes, No, NA
<b>kitchen</b>	This variable indicates the presence of a fitted kitchen. <b>kitcheng</b> is used to group kitchen as follows: Yes, No, NA
<b>eww</b>	If the warm water consumption was included in the energy consumption value calculation. <b>ewwg</b> variable is used to group eww as follows: Yes, No, NA
<b>subh</b>	It indicates whether a certificate of eligibility to public housing is needed to rent the apartment. <b>subhg</b> is used to group subh as follows: Yes, No, NA
<b>gtoilet</b>	This indicates the presence of a guest toilet. <b>gtoiletg</b> is used to group gtoilet as follows: Yes, No, NA
<b>garden</b>	This indicates the presence of a garden. <b>gardeng</b> is used to group garden as follows: Yes, No, NA
<b>hww</b>	If the warm water consumption was included in the heating cost value calculation. <b>hwwg</b> is used to group hww as follows: Yes, No, NA

## Continued

---

<b>cellar</b>	This indicates whether a property has a cellar room. <b>cellarg</b> is used to group cellar as follows: Yes, No, NA
<b>parking</b>	This variable indicates whether a parking space is available. <b>parking</b> is used to group parking as follows: Yes, No, NA
<b>furnishing</b>	This is an artificial category number indicating the property's facilities. <b>furnishingg</b> is used to group furnishing as follows: (Upscale, Luxury) = Upscale, (Normal, Simple) = Normal, no specification = NA
<b>eef</b>	This indicates the energy efficiency rating. <b>eefg</b> is used to group eeff as follows: (A, APLUS, B) = High, (C, D, E) = Medium, (F, G, H) = Low, no specification = NA
<b>ecert</b>	The type of energy performance certificate that the customer has for the object <b>ecertg</b> is used to group ecert as follows: Final energy demand = building, Energy consumption characteristic = consumption, NA
<b>pets</b>	This indicates whether pets are allowed in the property. <b>petsg</b> is used to group pets as follows: (Yes, by Agreement) = Yes, No = No, no specification = NA
<b>heat</b>	This indicates the type of heating. <b>heatg</b> is used to group heat as follows: Central Heating (CH), Non Central Heating (NCH), NA
<b>apcat</b>	This variable categorizes the property into different classes. <b>apcatg</b> is used to group apcat as follows: (Penthouse, Maisonette, Attic Apartment) = top, Apartment = middle, (Mezzanine, Terrace apartment) = low, Basement = below, NA
<b>pcon</b>	This indicates the condition of a property. <b>pcong</b> is used to group pcon as follows: (First occupancy, First occupancy after renovation) = First, (Maintained, as good as new) = Mt, In need of renovation = Inr, (Modernized, Renovated, Fully Renovated) = Md, NA

---

### 3.1. Data Sets

We split the data set described in **Table 1** and **Table 2** into two sub-data sets: Munich 2015 and Munich 2019. The number of rental properties contained in each data set is given in **Table 3**. The summaries of the response variable and the quantitative covariates are given in **Table 4** while in **Table 5**, we give the summary of each qualitative variable followed by their percentages.

**Table 3.** Number of rental properties in the two data sets.

City	2015	2019
Munich	14,449	14,776

**Table 4.** Univariate data summaries of quantitative covariates: first row = Munich 2015, second row = Munich 2019.

Variable	Summary						
	Min	25%	50%	Mean	75%	Max	NA
rent_sqm 2015	3.00	12.00	13.00	12.91	15.00	17.00	0
rent_sqm 2019	4.00	16.00	18.00	18.32	21.00	28.00	0
addcost							
	0.00	107.00	153.00	164.04	210.00	540.00	1355
	0.00	120.00	170.00	175.47	220.00	550.00	533
heatcost							
	0.00	60.00	85.00	89.35	110.00	288.00	10,075
	0.00	55.00	80.00	84.84	109.00	300.00	11,155
conyear							
	1860	1962	1976	1976	1999	2017	3522
	1858	1965	1985	1982	2014	2020	3068
lmod							
	1981	2011	2014	2012	2015	2016	9313
	1983	2013	2015	2014	2017	2018	11,386
lspace							
	23.00	55.00	71.00	73.79	90.00	161.00	0
	19.00	51.00	67.00	68.54	84.00	157.00	0
fspace							
	0.00	10.00	55.00	53.40	81.00	234.00	9276
	0.00	11.00	55.00	53.14	82.00	249.00	11,483
energycon							
	0.00	85.00	122.00	122.53	155.00	338.00	5975
	0.00	64.00	103.00	104.11	137.00	339.00	7379
adlength							
	0.00	0.00	0.00	0.58	1.00	20.00	0
	0.00	0.00	0.00	0.53	1.00	20.00	0

**Table 5.** Univariate data summaries of qualitative covariates: first row = Munich 2015, second row = Munich 2019.

Variable		Categories				
<b>afloorg</b>	(-1) - 0	1 - 2	3 - 9	>9	NA	
	1762 (0.12%)	6732 (0.47%)	3848 (0.27%)	63 (0%)	2044 (0.14%)	
	1648 (0.11%)	6428 (0.44%)	4687 (0.32%)	97 (0.01%)	1916 (0.13%)	
<b>bfloorg</b>	0 - 2	3	4	5	>5	NA
	2744 (0.19%)	2226 (0.15%)	2770 (0.19%)	1833 (0.13%)	1418 (0.1%)	3458 (0.24%)
	2741 (0.19%)	2187 (0.15%)	2517 (0.17%)	2160 (0.15%)	1950 (0.13%)	3221 (0.22%)
<b>nroomsg</b>	1 - 1.5	2 - 2.5	3 - 3.5	>3.5		
	2157 (0.15%)	5710 (0.4%)	4841 (0.34%)	1741 (0.12%)		
	2836 (0.19%)	5949 (0.4%)	4768 (0.32%)	1223 (0.08%)		
<b>nbedg</b>	0 - 1	2	>2	NA		
	5636 (0.39%)	3562 (0.25%)	1240 (0.09%)	4011 (0.28%)		
	3884 (0.26%)	2329 (0.16%)	684 (0.05%)	7879 (0.53%)		
<b>nbathg</b>	0 - 1	>1	NA			
	10,690 (0.74%)	1669 (0.12%)	2090 (0.14%)			
	11,310 (0.77%)	1657 (0.11%)	1809 (0.12%)			
<b>elevatorg</b>	Yes	No	NA			
	6125 (0.42%)	8108 (0.56%)	216 (0.01%)			
	7929 (0.54%)	6847 (0.46%)	0 (0%)			
<b>balconyg</b>	Yes	No	NA			
	10,863 (0.75%)	3406 (0.24%)	180 (0.01%)			
	11,554 (0.78%)	3222 (0.22%)	0 (0%)			

## Continued

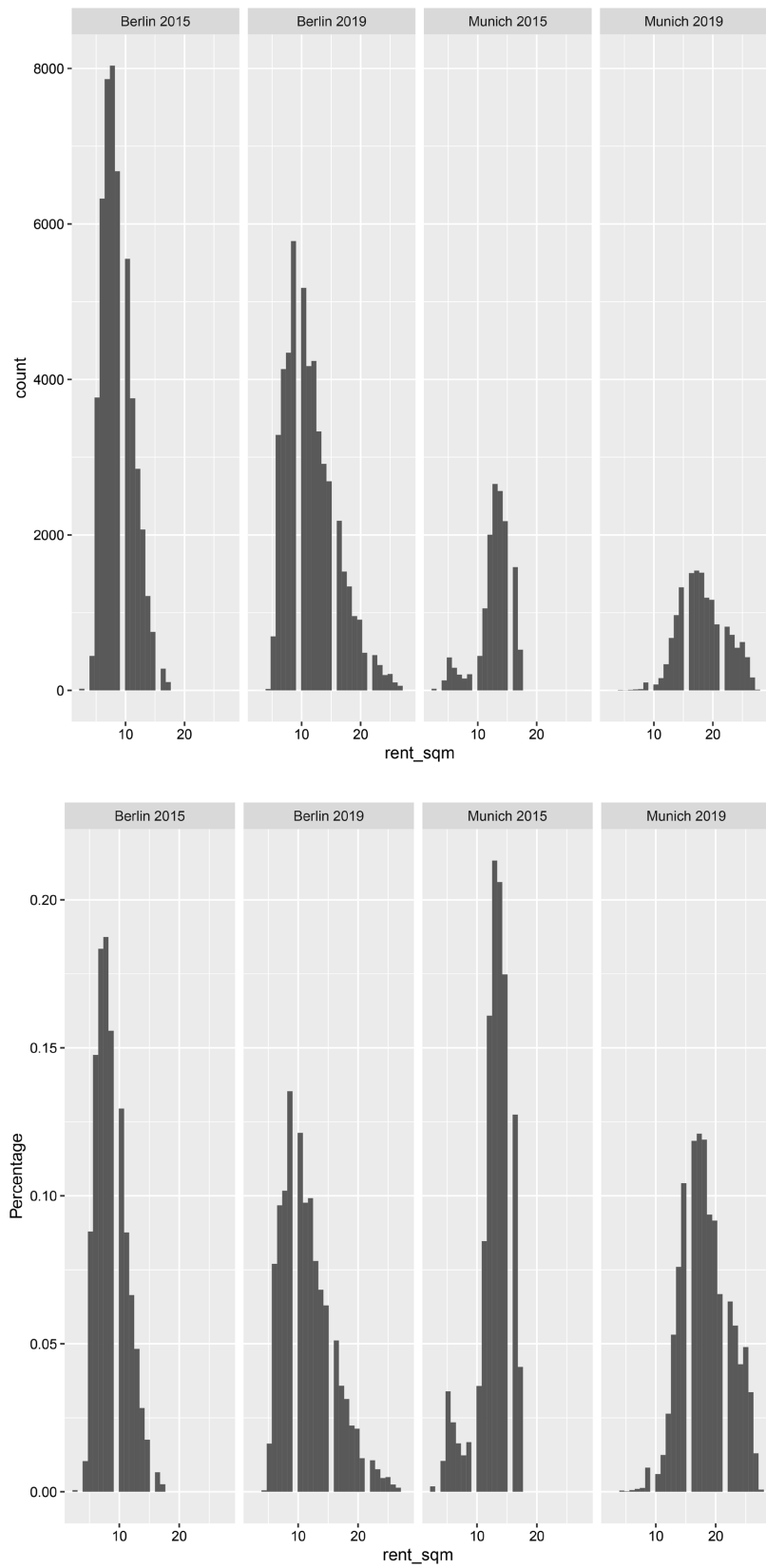
<b>kitcheng</b>	Yes	No	NA
	8756 (0.61%)	5438 (0.38%)	255 (0.02%)
	9878 (0.67%)	4898 (0.33%)	0 (0%)
<b>ewwg</b>	Yes	No	NA
	3775 (0.26%)	10,454 (0.72%)	220 (0.02%)
	1419 (0.1%)	723 (0.05%)	12,634 (0.86%)
<b>subhg</b>	Yes	No	NA
	30 (0.00%)	12,534 (0.87%)	1885 (0.13%)
	162 (0.01%)	14,614 (0.99%)	0 (0%)
<b>gtoiletg</b>	Yes	No	NA
	3186 (0.22%)	11,254 (0.78%)	9 (0.00%)
	2948 (0.20%)	11,828 (0.80%)	0 (0%)
<b>gardeng</b>	Yes	No	NA
	2726 (0.19%)	11,173 (0.77%)	550 (0.04%)
	3074 (0.21%)	11,702 (0.79%)	0 (0%)
<b>hwwg</b>	Yes	No	NA
	8856 (0.61%)	4320 (0.3%)	1273 (0.09%)
	10,161 (0.69%)	4088 (0.28%)	527 (0.04%)
<b>cellarg</b>	Yes	No	NA
	11,315 (0.78%)	3036 (0.21%)	98 (0.01%)
	11,533 (0.78%)	3243 (0.22%)	0 (0%)
<b>parkingg</b>	Yes	No	NA
	59 (0.00%)	0 (0%)	14,390 (1.00%)
	7911 (0.54%)	228 (0.02%)	6637 (0.45%)

Continued

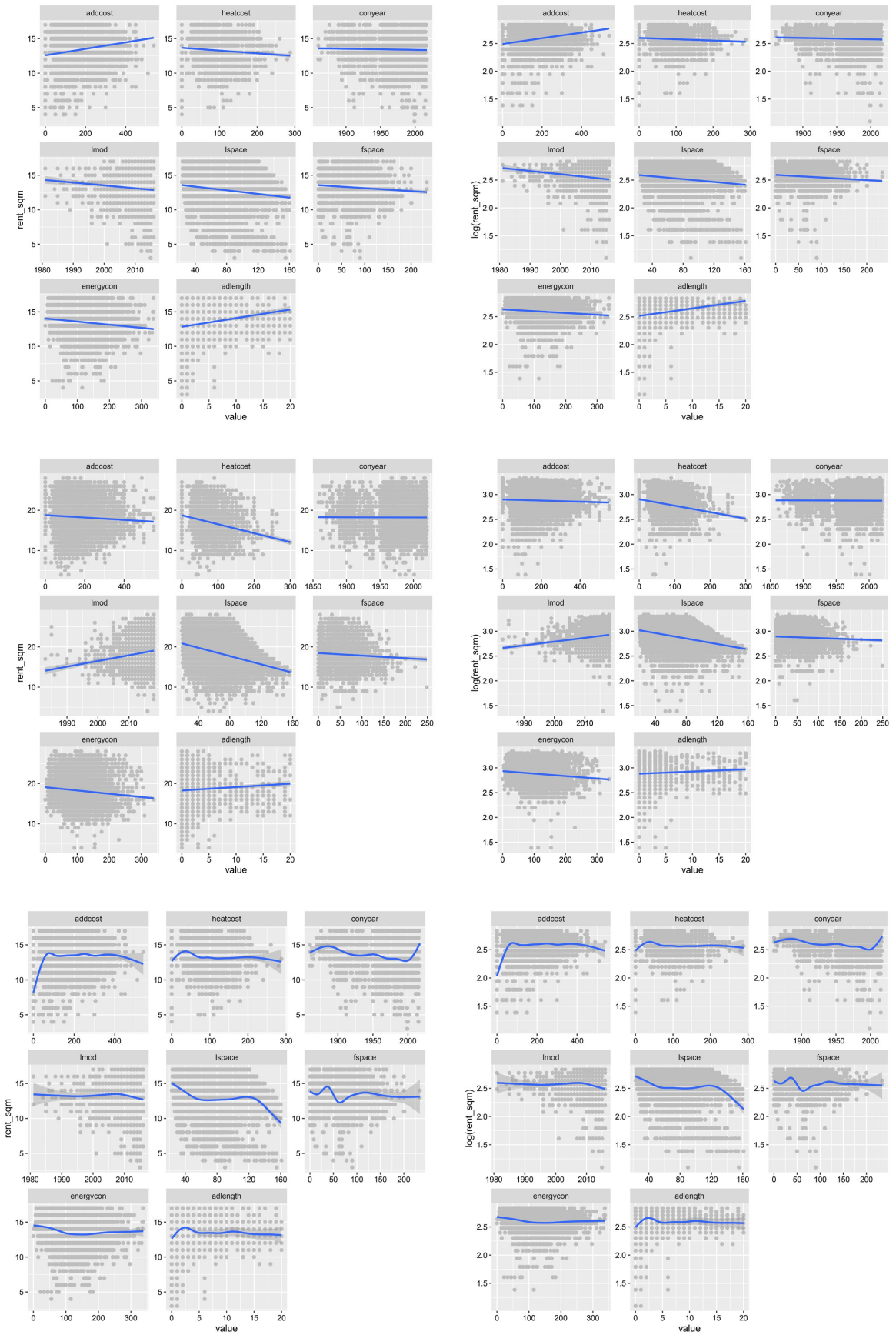
<b>furnishngg</b>	Upscale	Normal	NA		
	5699 (0.39%)	3591 (0.25%)	5159 (0.36%)		
	7156 (0.48%)	2726 (0.18%)	4894 (0.33%)		
<b>efffg</b>	High	Medium	Low	NA	
	314 (0.02%)	257 (0.02%)	63 (0%)	13,815 (0.96%)	
	474 (0.03%)	388 (0.03%)	50 (0%)	13,864 (0.94%)	
<b>ecertg</b>	building	consumption	NA		
	2898 (0.20%)	6027 (0.42%)	5524 (0.38%)		
	3228 (0.22%)	4393 (0.30%)	7155 (0.48%)		
<b>petsg</b>	Yes	No	NA		
	947 (0.07%)	4123 (0.29%)	9379 (0.65%)		
	3460 (0.23%)	5629 (0.38%)	5687 (0.38%)		
<b>heatg</b>	CH	NCH	NA		
	8056 (0.56%)	3744 (0.26%)	2649 (0.18%)		
	6589 (0.45%)	5560 (0.38%)	2627 (0.18%)		
<b>apcatg</b>	top	middle	low	below	NA
	2011 (0.14%)	7627 (0.53%)	515 (0.04%)	80 (0.01%)	4216 (0.29%)
	2066 (0.14%)	7977 (0.54%)	1158 (0.08%)	130 (0.01%)	3445 (0.23%)
<b>pcong</b>	First	Mt	Md	Inr	NA
	1781 (0.12%)	5525 (0.38%)	3280 (0.23%)	17 (0%)	3846 (0.27%)
	2682 (0.18%)	5537 (0.37%)	3175 (0.21%)	11 (0%)	3371 (0.23%)

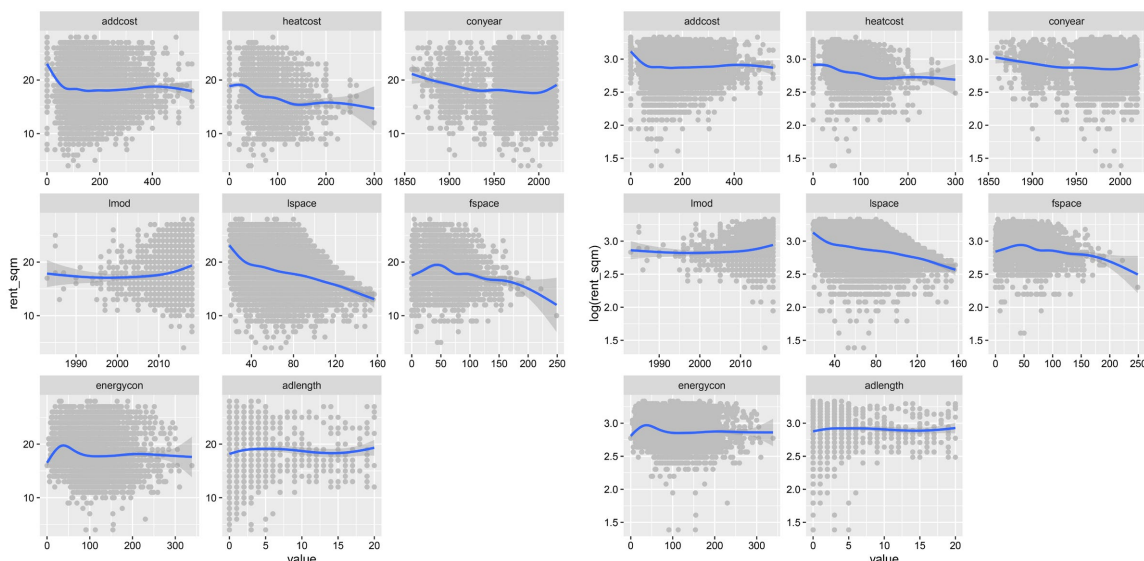
### 3.2. Exploratory Data Analysis (EDA)

See **Figures 1-3**.

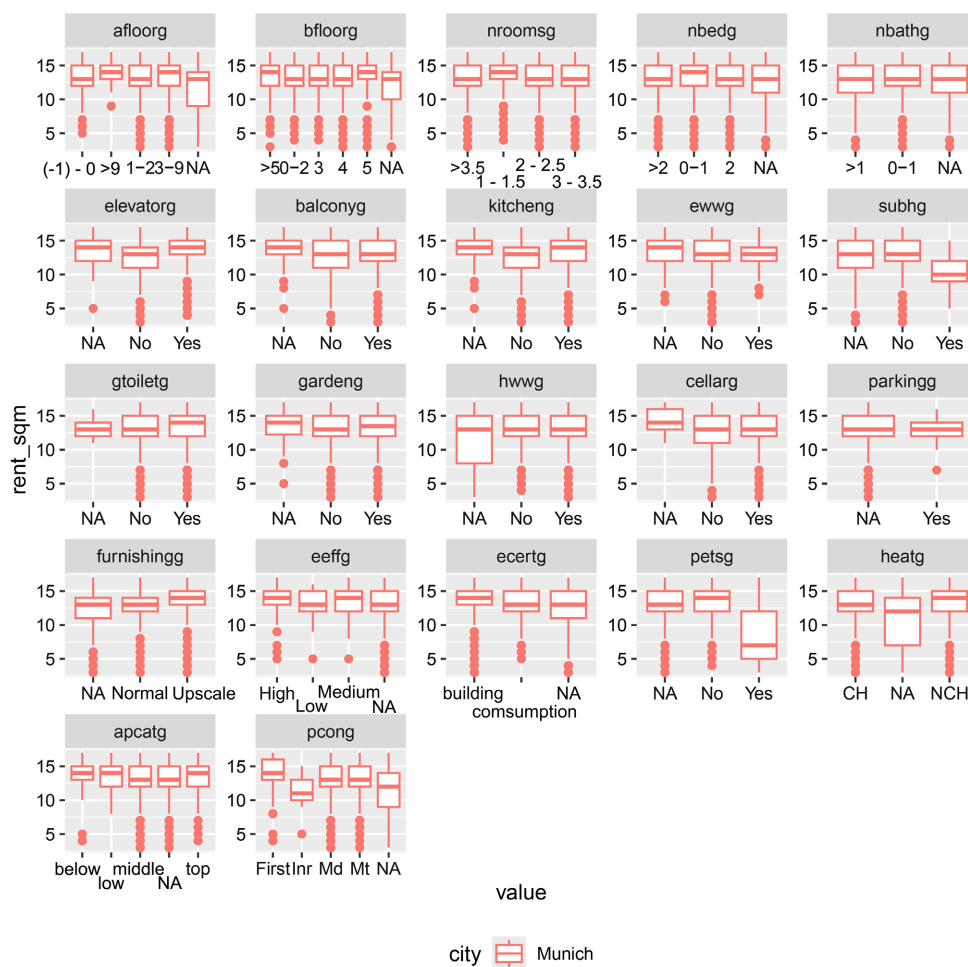


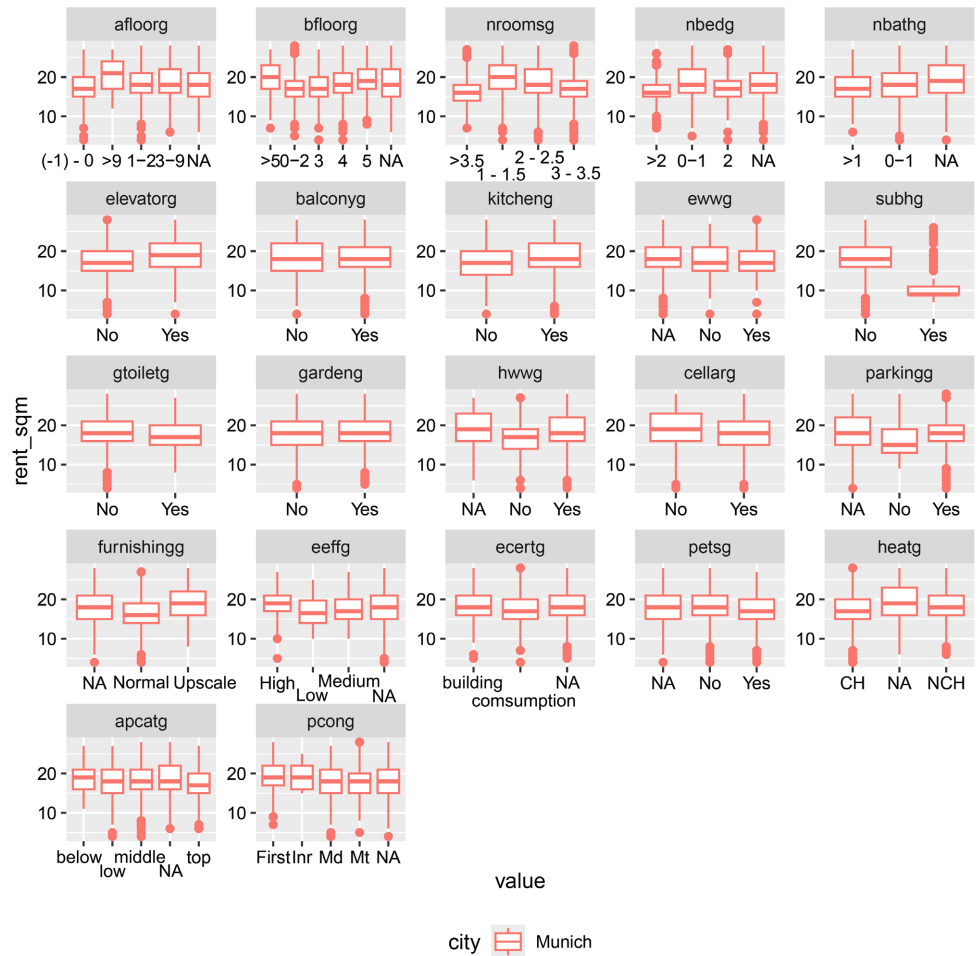
**Figure 1.** Histograms of response variable—**rent\_sqm**: first column = **counts**, second column = **percentage**.





**Figure 2.** Scatter plots of quantitative covariates versus response ( $rent\_sqm$ ) with **Linear Smooth (LS)** and **Non Linear Smooth (NLS)**: first column = ( $rent\_sqm$ ) and second column =  $\log(rent\_sqm)$ . (first row) = Munich 2015 with LS, (second row) = Munich 2019 with LS, (third row) = Munich 2015 with NLS, (fourth row) = Munich 2019 with NLS.





**Figure 3.** Box plots of qualitative covariates versus response (rent\_sqm): first column = **Munich 2015**, second column = **Munich 2019**.

### 3.3. Interpretation of Main Effects for the Quantitative and Qualitative Covariates

Looking at the above transformations on rent\_sqm in **Table 6** and **Table 7**, we may likely go with the log transformation for linear and non-linear covariates based on its suitability with respect to constant variance discussed in Section 2 and the effects of the covariates on rent\_sqm.

**Table 6.** Interpretation of main effects for the quantitative covariates on **rent\_sqm** and **log(rent\_sqm)** in **Figure 2**: first block = **Linear smooth**, second block = **Nonlinear smooth**.

Variables	Munich 2015 (rent_sqm)	Munich 2019 (rent_sqm)	Munich 2015 (log(rent_sqm))	Munich 2019 (log(rent_sqm))
<b>Addcost</b>	Linear (increasing)	Nearly constant	Linear (increasing)	Nearly constant
<b>Heatcost</b>	Nearly constant	Linear (decreasing)	Nearly constant	Linear (decreasing)
<b>Conyear</b>	Constant	Constant	Constant	Constant

## Continued

<b>Lmod</b>	Nearly constant	Linear (increasing)	Linear (decreasing)	Linear (increasing)
<b>Lspace</b>	Linear (decreasing)	Linear (decreasing)	Nearly constant	Linear (decreasing)
<b>Fspace</b>	Nearly constant	Nearly constant	Nearly constant	Constant
<b>Energycon</b>	Nearly constant	Linear (decreasing)	Nearly constant	Nearly constant
<b>Adlength</b>	Linear (increasing)	Linear (increasing)	Linear (increasing)	Nearly constant
<b>Addcost</b>	Quadratic	Quadratic	Quadratic	Quadratic
<b>Heatcost</b>	Quadratic	Quadratic	Nearly linear	Nearly linear
<b>Conyear</b>	Cubic	Quadratic	Quadratic	Nearly linear
<b>Lmod</b>	Nearly linear	Nearly constant	Nearly linear	Nearly constant
<b>lspace</b>	Cubic	Nearly linear	Cubic	Linear (decreasing)
<b>fspace</b>	Cubic	Quadratic	Cubic	Quadratic
<b>Energycon</b>	Quadratic	Quadratic	Nearly constant	Quadratic
<b>Adlength</b>	Quadratic	Nearly constant	Quadratic	Constant

**Table 7.** Interpretation of main effects for the qualitative covariates on **rent\_sqm** in **Munich 2015** and **Munich 2019** in **Figure 3**.

<b>Variables</b>	<b>Munich 2015</b>	<b>Munich 2019</b>
<b>afloorg</b>	No	Yes
<b>bfloorg</b>	Yes	Yes
<b>nroomsg</b>	Yes	Yes
<b>nbedg</b>	No	Yes
<b>nbathg</b>	No	No
<b>elevatorg</b>	Yes	Yes
<b>balconyg</b>	No	No
<b>kitcheng</b>	Yes	Yes
<b>ewwg</b>	No	No
<b>subhg</b>	Yes	Yes
<b>gtoiletg</b>	No	No
<b>gardeng</b>	No	No
<b>hwwg</b>	No	Yes
<b>cellarg</b>	Yes	Yes
<b>parkingg</b>	No	Yes
<b>furnishingg</b>	Yes	Yes

Continued

<b>effgg</b>	Yes	Yes
<b>ecertg</b>	No	No
<b>petsg</b>	Yes	No
<b>heatg</b>	Yes	Yes
<b>apcatg</b>	No	No
<b>pcong</b>	Yes	No

### 4. Model Fittings and Predictions

We discuss how we select the type of model we use to fit the rent\_sqm for Munich rental properties in 2015 and 2019. To refine the regression model for the rent per square meter in Munich, a stepwise backward regression was applied using the step() function in R. This method began with a full model containing all relevant predictors and iteratively removed nonsignificant variables based on the Akaike Information Criterion (AIC). The backward selection process ensured a more parsimonious model by retaining only the most influential variables, enhancing interpretability while maintaining predictive strength and minimizing model complexity. We first fit four models for the response variable in Munich 2015 in the following cases:

- **Case 1:** We fit a linear regression model where we do not transform the response variable against the covariates ( $\text{lm}(\text{rent\_sqm} \sim \text{addcost} + \text{heatcost} + \text{conyear} + \dots + \text{pcong}, \text{data} = \text{dm5\_fit})$ ).
- **Case 2:** We fit the log of the response variable against the covariates ( $\text{lm}(\log(\text{rent\_sqm}) \sim \text{addcost} + \text{heatcost} + \text{conyear} + \dots + \text{pcong}, \text{data} = \text{dm5\_fit})$ ).
- **Case 3:** We include a non-linear covariates against the response variable ( $\text{lm}(\text{rent\_sqm} \sim \text{poly}(\text{addcost}, 2) + \text{heatcost} + \text{poly}(\text{conyear}, 3) + \dots + \text{pcong}, \text{data} = \text{dm5\_fit})$ ).
- **Case 4:** We include a non-linear covariates against the log of the response variable ( $\text{lm}(\log(\text{rent\_sqm}) \sim \text{poly}(\text{addcost}, 2) + \text{heatcost} + \text{poly}(\text{conyear}, 3) + \dots + \text{pcong}, \text{data} = \text{dm5\_fit})$ ).

We also do similar model fitting (the 4 cases) for Munich 2019. The summaries are found in **Table 8**.

**Table 8.** Model fitting summary with only main effect.

Munich 2015	Case 1	Case 2	Case 3	Case 4
Adjusted R-square	0.2762	0.2652	0.3101	0.2879
Number of parameters ( <i>p</i> )	38	38	39	33
Munich 2019				
Adjusted R-square	0.5139	0.5145	0.3078	0.5468
Number of parameters ( <i>p</i> )	25	22	41	27

Looking at the model fitting summary in **Table 8**, we decided to go with case 4, which is the log transformation on rent\_sqm ( $\log(\text{rent\_sqm})$ ) for the non-linear covariates as it relatively satisfied most of the listed assumptions with a larger R-square, compared to the others in the four data sets.

#### 4.1. Model Fitting of Log(Rent\_Sqm) on Non-Linear Covariates for Munich 2015 and Munich 2019

See **Table 9** and **Table 10**.

**Table 9.** Munich 2015.

	Estimate	Std. Error	t value	Pr (> t )
(Intercept)	2.5838	0.0740	34.90	0.0000
poly (conyear, 2) 1	-0.5052	0.1259	-4.01	0.0001
poly (conyear, 2) 2	0.5825	0.1307	4.46	0.0000
poly (lspace, 3) 1	-1.4841	0.2413	-6.15	0.0000
poly (lspace, 3) 2	0.2820	0.1631	1.73	0.0842
poly (lspace, 3) 3	-0.3785	0.1423	-2.66	0.0080
adlength	0.0066	0.0030	2.19	0.0290
nroomsg 1 - 1.5	-0.0973	0.0317	-3.06	0.0023
nroomsg 2 - 2.5	-0.0877	0.0227	-3.87	0.0001
nroomsg 3 - 3.5	-0.0542	0.0182	-2.97	0.0031
nbedg 0 - 1	0.0638	0.0211	3.02	0.0026
nbedg 2	0.0478	0.0197	2.42	0.0157
nbedgNA	0.0977	0.0279	3.50	0.0005
elevatorgYes	0.0378	0.0096	3.93	0.0001
kitchengNo	-0.0606	0.0383	-1.58	0.1141
kitchengYes	-0.0272	0.0395	-0.69	0.4909
ewwgNo	-0.0808	0.0449	-1.80	0.0721
ewwgYes	-0.0976	0.0450	-2.17	0.0304
subhgNo	0.0422	0.0219	1.93	0.0545
gtoiletgYes	0.0268	0.0138	1.94	0.0528
hwwgYes	0.0205	0.0104	1.98	0.0483
furnishinggNormal	-0.0034	0.0185	-0.18	0.8562
furnishinggUpscale	0.0768	0.0182	4.23	0.0000

## Continued

eeffgLow	0.1780	0.0658	2.70	0.0070
eeffgMedium	0.1156	0.0478	2.42	0.0158
eeffgNA	0.0725	0.0447	1.62	0.1055
petsgNo	-0.0098	0.0103	-0.95	0.3441
petsgYes	-0.0651	0.0316	-2.06	0.0399
heatgNA	0.0706	0.0213	3.31	0.0010
heatgNCH	-0.0052	0.0121	-0.43	0.6675
pcongInr	0.0700	0.1188	0.59	0.5561
pcongMd	-0.0529	0.0156	-3.40	0.0007
pcongMt	-0.0530	0.0161	-3.30	0.0010
pcongNA	0.0270	0.0262	1.03	0.3031
Observations	711			
R <sup>2</sup>	0.321			
Adj. R <sup>2</sup>	0.288			
Residual Std. Error	0.116 (df = 677)			
F Statistic	9.698*** (df = 33; 677)			
p-value	<2.2e-16			

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

**Table 10.** Munich 2019.

	Estimate	Std. Error	t value	Pr (> t )
(Intercept)	-8.9984	4.7335	-1.90	0.0586
heatcost	0.0004	0.0003	1.47	0.1425
lmod	0.0060	0.0023	2.54	0.0117
poly (lspace, 2) 1	-1.2984	0.2104	-6.17	0.0000
poly (lspace, 2) 2	0.4809	0.1481	3.25	0.0013
poly (fspace, 2) 1	-0.1854	0.1688	-1.10	0.2733
poly (fspace, 2) 2	-0.4037	0.1690	-2.39	0.0178
poly (energycon, 2) 1	0.0216	0.1557	0.14	0.8901
poly (energycon, 2) 2	0.4404	0.1571	2.80	0.0055
bfloorg 0 - 2	-0.0579	0.0340	-1.70	0.0903

## Continued

bfloorg 3	-0.0029	0.0335	-0.09	0.9322
bfloorg 4	-0.0181	0.0314	-0.58	0.5648
bfloorg 5	0.0491	0.0320	1.53	0.1266
bfloorgNA	-0.0580	0.1068	-0.54	0.5877
kitchengYes	0.0822	0.0230	3.57	0.0004
hwwgYes	0.0551	0.0216	2.55	0.0116
parkinggNo	-0.1027	0.0561	-1.83	0.0686
parkinggYes	-0.0336	0.0198	-1.69	0.0918
furnishinggNormal	-0.0400	0.0398	-1.01	0.3159
furnishinggUpscale	0.1132	0.0385	2.94	0.0036
ecertgconsumption	-0.0471	0.0214	-2.20	0.0288
apcatgLow	-0.1004	0.1636	-0.61	0.5400
apcatgMiddle	-0.1517	0.1564	-0.97	0.3332
apcatgNA	-0.2236	0.1587	-1.41	0.1604
apcatgTop	-0.0951	0.1565	-0.61	0.5441
pcongMd	-0.1170	0.0445	-2.63	0.0091
pcongMt	-0.1001	0.0460	-2.18	0.0305
pcongNA	-0.0194	0.0576	-0.34	0.7362
Observations	244			
R <sup>2</sup>	0.597			
Adj. R <sup>2</sup>	0.547			
Residual Std. Error	0.137 (df = 216)			
F Statistic	11.859*** (df = 27; 216)			
p-value	<2.2e-16			

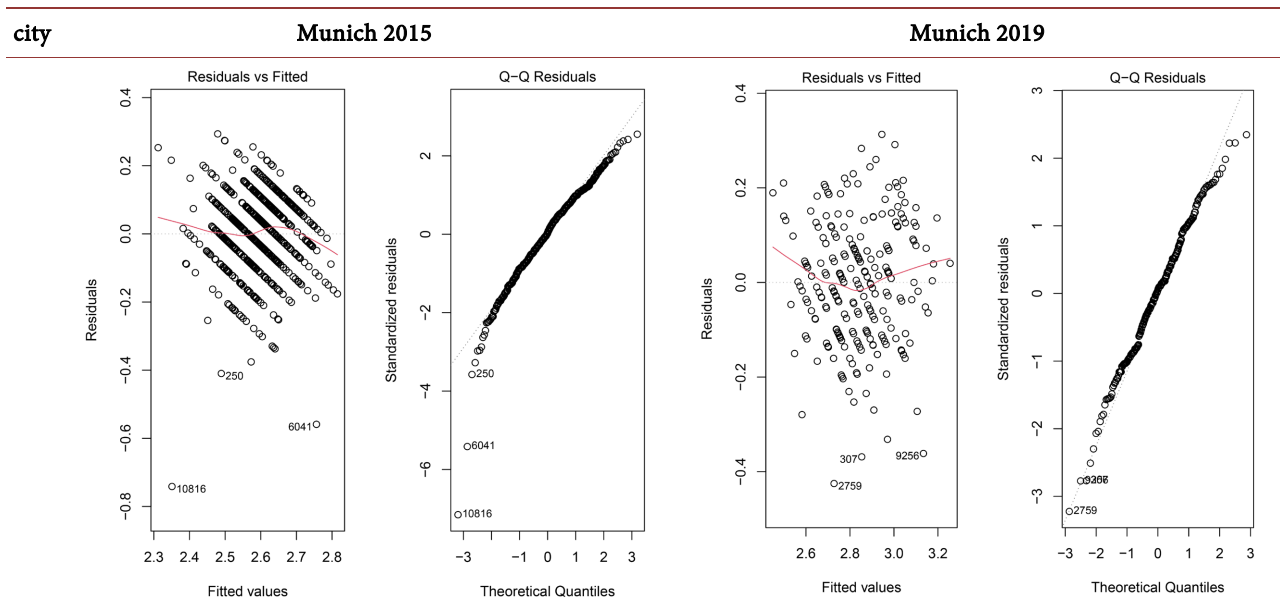
$p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ .

## 4.2. Residual Plots of Model Fittings

We plot the residuals versus the fitted values to see if there is a trend to check for the plausibility of the linearity assumption discussed in Section 2. Also, we plot the QQ plots of the covariates versus the theoretical normal quantile to see if it is a straight line to check for the plausibility of the normality assumption, which was discussed in Section 2.

From the plots in **Table 11**, we find that the fitted models do not relatively violate the linear regression assumptions in Section 2.19.

**Table 11.** Residual plots of model fittings for **Munich 2015** and **Munich 2019**.



### 4.3. Model Predictions of Rent\_Sqm for the Main Effect Models

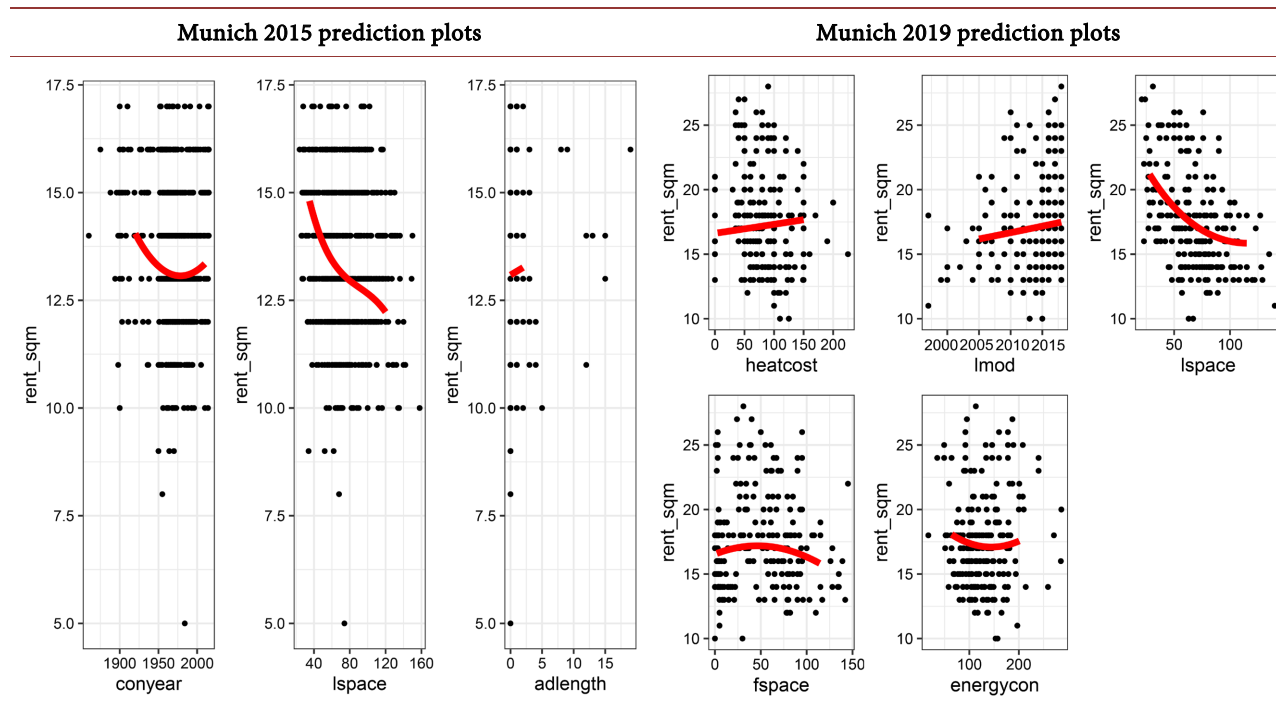
In this section, we will predict the values of rent\_sqm for the main effect models given in **Table 9** and **Table 10** using the most influential variables from the pairwise selection as shown in **Table 12** and **Table 13**. We will use the median of the continuous covariates and the mode of the qualitative covariates for our prediction. We consider the mode for the qualitative covariates and the median for the remaining continuous variables while we take 50 values between the 5th and 95th quantile/percentile of the variable we are plotting. We also consider the different categories of each qualitative covariate which we are using for the prediction of rent\_sqm while other qualitative covariates remain in their mode and the continuous covariates in their medians respectively. We also computed the Variance Inflation Factor (VIF) for all predictors. The GVIF and adjusted GVIF<sup>1/(2 - Df)</sup> values were all below 2, as shown in **Table 14**, indicating no significant multicollinearity issues. This implies that the predictors were sufficiently independent of each other. Thus, no predictor variables were removed on this basis, and the model structure remains statistically robust.

## 5. Findings

### 5.1. Summary of Findings

In **Figure 1**, there is a significant shift in the histogram plots of rent\_sqm for Munich 2015 and Munich 2019. For instance, in Munich 2015, we can see that the rent\_sqm is below 20 Euros, but in 2019, the rent\_sqm is over 20 Euros. This shows that the rent price increases with time, which is also confirmed in our prediction. For instance, the predicted rent\_sqm increased in Munich from 2015 to 2019 by 31.17%, 31.17%, and 39.86% with apartments that have a kitchen, Upscale furnishing, and First occupancy condition.

**Table 12.** Model predictions of **rent\_sqm** for the influential **quantitative covariates**.



**Table 13.** Munich 2015.

Variables	categories	Munich 2015	Munich 2019
<b>afloorg</b>	(-1) - 0		
	1 - 2		
	3 - 9		
	>9		
	NA		
<b>bfloorg</b>	0 - 2		16.50
	3		17.44
	4		17.17 (mode = 4)
	5		18.37
	>5		17.49
	NA		16.50
<b>nroomsg</b>	1 - 1.5	12.96	
	2 - 2.5	13.09 (mode = 2 - 2.5)	
	3 - 3.5	13.53	
	>3.5	14.29	

Continued

	0 - 1	13.09 (mode = 0 - 1)	
<b>nbedg</b>	2	12.88	
	>2	12.28	
	NA	13.54	
	0 - 1		
<b>nbathg</b>	>1		
	NA		
	Yes	13.59	
<b>elevatorg</b>	No	13.09 (mode = No)	
	NA		
<b>balconyg</b>			
	Yes	13.09 (mode = Yes)	17.17 (mode = Yes)
<b>kitcheng</b>	No	12.66	15.82
	NA	13.45	
	Yes	12.87	
<b>ewwg</b>	No	13.09 (mode = No)	
	NA	14.19	
	Yes		
<b>subhg</b>	No	13.09 (mode = No)	
	NA	12.55	
	Yes	13.44	
<b>gtoiletg</b>	No	13.09 (mode = No)	
	NA		
	Yes		
<b>gardeng</b>	No		
	NA		
	Yes	13.36	18.14
<b>hwwg</b>	No	13.09 (mode = No)	17.17 (mode = No)
	NA		
	Yes		
<b>cellarg</b>	No		
	NA		
	Yes	17.17 (mode = Yes)	
<b>parkingg</b>	No	16.02	
	NA	17.77	
	Upscale	13.09 (mode = Upscale)	17.17 (mode = Upscale)
<b>furnishingg</b>	Normal	12.08	14.73
	NA	12.12	15.33

## Continued

<b>eeffgg</b>	High	12.17	
	Meduim	13.66	
	Low	14.54	
	NA	13.087 (mode = NA)	
<b>ecertg</b>	consumption		17.17 (mode = consumption)
	building		18.00
<b>petsg</b>	Yes	12.26	
	No	12.96	
	NA	13.09 (mode = NA)	
<b>heatg</b>	CH	13.09 (mode = CH)	
	NCH	13.02	
	NA	14.04	
<b>apcatg</b>	top		18.17
	middle		17.17 (mode = middle)
	low		18.08
	below		19.99
	NA		15.98
<b>pcong</b>	Md	13.09 (mode = Md)	17.17 (mode = Md)
	Mt	13.09	17.46
	First	13.80	19.30
	Inr	14.80	
	NA	14.18	18.93

Table 14. VIF Munich 2019.

	GVIF	Df	$GVIF^{1/(2 * Df)}$
heatcost	1.68	1.00	1.30
lmod	1.37	1.00	1.17
poly (lspace, 2)	2.76	2.00	1.29
poly (fspace, 2)	2.25	2.00	1.23
poly (energycon, 2)	1.69	2.00	1.14
bfloorg	2.18	5.00	1.08
kitcheng	1.41	1.00	1.19
hwwg	1.25	1.00	1.12
parkingg	1.39	2.00	1.09
furnishingg	1.50	2.00	1.11
ecertg	1.24	1.00	1.11
apcatg	2.03	4.00	1.09
pcong	1.81	3.00	1.10

From **Table 12**, we can summarise the behaviour of the **predicted rent\_sqm** for the influential quantitative covariates as follows:

- In **Munich 2015**, all the variables have an influence on **rent\_sqm**. The length of advertisement enters the model linearly and has an increasing trend with the **rent\_sqm** while the construction year and living space variables enter the model nonlinearly, although we see a decreasing trend in the living space with the **rent\_sqm**.
- Also, in **Munich 2019**, all the variables equally have an influence on **rent\_sqm**. The heat cost and the last modernization variables enter the model linearly and have an increasing trend with the **rent\_sqm** while the other variables enter the model nonlinearly, although we see a decreasing trend in the living space with the **rent\_sqm**.

In **Table 13**, we can summarise the behaviour of the predicted rent\_sqm for the influential qualitative covariates as follows:

- The predicted **rent\_sqm** is highest in Munich, estimated at 19.99 euros in 2019.
- **Building floors (bfloor)**: With 5 building floors apartments, our predicted **rent\_sqm** is at the highest 18.37 Euros for Munich 2019. Also, the predicted **rent\_sqm** is at the lowest (16.50 Euros) with 0 - 2 building floors apartments.
- **Number of rooms (nrooms)**: In **Munich 2015**, our predicted rent\_sqm value is at the highest (14.29 Euros) with apartments that have **>3.5 rooms** while it is at the lowest (12.96 Euros) with apartments that have 1 - 1.5 rooms.
- **Number of bedrooms (nbed)**: In **Munich 2015**, we can see a **decreasing trend** in the **predicted rent\_sqm** (13.09, 12.88, and 12.28 Euros) with respect to the same order of the categories of the **number of bedrooms (0 - 1, 2, and >2)**. Thus rent\_sqm seems to decrease in Munich with apartments that have a lower number of bedrooms.
- **Elevator**: We can see an **increase** in the **predicted rent\_sqm** (13.59 Euros) for apartments with an elevator in Munich 2015, unlike the apartments without an elevator, where the **predicted rent\_sqm** is (13.09 Euros). Thus, rent\_sqm seems to increase with apartments that have an elevator (vice versa).
- **Kitchen**: We can also see an **increase** in the **predicted rent\_sqm** (13.09, and 17.17 Euros) for apartments with a kitchen in Munich 2015 and Munich 2019 respectively unlike the apartments without a kitchen where the **predicted rent\_sqm** are respectively (12.66, and 15.82 Euros). Thus rent\_sqm seems to increase in Munich with apartments that have a kitchen (vice versa).
- **Eww**: The **predicted rent\_sqm** in Munich 2019 is lower (12.87 Euros) with apartments that have the inclusion of warm water consumption in the energy consumption value calculation compared with the apartments that do not have it (13.09 Euros). Thus rent\_sqm seems to decrease in Munich with apartments that have the inclusion of warm water consumption in the energy consumption value calculation (vice versa).
- **Gtoilet**: The **predicted rent\_sqm** is higher with apartments that have a guest toilet (13.44 Euros) in Munich 2015, compared to the apartments with no guest

toilet (13.08 Euros). Thus `rent_sqm` seems to increase with apartments that have a guest toilet in Munich (vice versa).

- **Hww:** With apartments that have the warm water consumption included in the heating cost value calculation in both Munich 2015 and 2019, the **predicted `rent_sqm`** is higher (13.36 and 18.14 Euros) compared to apartments that do not have it (13.09 and 17.17 Euros), thereby increased by 36% in Munich from 2015 to 2019 with apartments that have the warm water consumption included in the heating cost value calculation. Thus, `rent_sqm` seems to increase with apartments that have the warm water consumption included in the heating cost value calculation in Munich (vice versa).
- **Parking space:** In Munich 2019, the **predicted `rent_sqm`** is also higher (17.17 Euros) for apartments that have a parking space compared to apartments that do not have a parking space (16.02 Euros). Thus `rent_sqm` seems to increase in Munich with apartments that have a parking space (vice versa).
- **Furnishing:** The predicted `rent_sqm` is at the highest with apartments that have **Upscale furnishing** for **Munich 2015**, and **Munich 2019**. Also, with Upscale furnished apartments, the predicted `rent_sqm` in Munich increased from 2015 to 2019 by 31.17%. It equally increased from Normal to Upscale furnishing apartments by 8.36% and 16.56% for Munich 2015 and Munich 2019 respectively. Thus, `rent_sqm` seems to increase with apartments that have Upscale furnishing in Munich (vice versa), as well as with respect to time.
- **Energy efficiency rating (eeff):** We can also see a **decreasing trend** in the **predicted `rent_sqm`** with respect to the order of the categories of **energy efficiency rating (Low, Medium, and High)** (14.54, 13.67, and 12.17 Euros). Thus `rent_sqm` seems to decrease in Munich with respect to the order of energy efficiency rating categories (Low, Medium, and High) (vice versa).
- **Ecertg:** In Munich 2019, the **predicted `rent_sqm`** is higher with apartments that have the **building type of energy performance certificate** (18.00 Euros) compared to apartments that have the **construction type of energy performance certificate** (17.17 Euros).
- **Pets:** The **predicted `rent_sqm`** is lower with apartments that allow pets (12.26 Euros) in Munich 2015, compared to the apartments that do not allow pets (12.96 Euros), thereby decreasing by 5% for Munich 2015. Thus, `rent_sqm` seems to increase with apartments that do not allow pets in Munich (vice versa).
- **Heat:** Our predicted `rent_sqm` is higher with apartments that make use of the **central heating (CH)** as their heating type (13.09 Euros) in Munich 2015, compared to the apartments that make use of the **Non-Central Heating (NCH)** as their heating type (13.02 Euros), thereby increasing by 1% for Munich 2015. Thus, `rent_sqm` seems to increase with apartments that make use of the central heating as their heating type in Munich (vice versa).
- **Apartment categories (apcat):** Our predicted `rent_sqm` is at the highest with the **below category** apartments (19.99 Euros) for Munich 2019, but it is at the lowest with the **middle category** apartments (17.17 Euros).

- **Property condition categories (pcon):** Our predicted **rent\_sqm** is at the highest with the **First occupancy condition** apartments (19.30 Euros) for **Munich 2019**, but it is at the highest with the **in need of renovation condition** apartments (14.80 Euros) for Munich 2015. Thus, **rent\_sqm** is relatively higher for the first occupancy condition apartments in Munich compared to other apartment condition categories.

## 5.2. Discussion on Research Questions

### 5.2.1. RQ1: Is There Any Relationship between the Response Variable (Rent Per Sqm) and the Predictors?

The analysis indicates significant relationships between rent per square meter and various predictors. In both the Munich 2015 and 2019 datasets, all examined variables influenced rent per sqm. For instance, in Munich 2015, the advertisement length showed a linear and increasing relationship with rent per sqm, while construction year and living space exhibited nonlinear relationships. Similarly, in Munich 2019, heat cost and last modernization had linear increasing trends, whereas other variables, including living space, displayed nonlinear associations.

These findings align with broader market trends. According to JLL's Housing Market Overview for H2 2023, Munich remains Germany's most expensive housing market, with asking rents rising by 5.1% year-on-year to €22.50/sqm per month. This suggests that various factors, including those studied, contribute to rent variations [26].

### 5.2.2. RQ2: Does the Relationship between the Response Variable and the Predictors Require a Transformation to Satisfy Linear Regression Assumptions?

A log transformation was applied to the rent per sqm variable to address potential non-linear relationships and meet linear regression assumptions. This transformation improved the model's fit, as evidenced by a higher R-squared value, indicating a better explanation of variance in the response variable. Such transformations are commonly employed in housing market analyses to stabilize variance and normalize distributions [4]. This approach is consistent with standard econometric modeling practices in real estate studies, where log-linear models account for skewness and heteroscedasticity in rent and housing price distributions [3]. This transformation approach also supports previous findings in [27]-[29].

### 5.2.3. RQ3: What Are the Key Predictors (Covariates) That Significantly Influence the Rental Price Per Square Meter in Munich's Housing Market?

The study identified several key predictors impacting rental prices:

- **Quantitative Covariates:** In Munich 2015, the advertisement length had a linear and positive relationship with rent per sqm, while construction year and living space exhibited non-linear effects. In 2019, heat cost and last modernization showed linear positive trends, with other variables displaying non-linear relationships.

- **Qualitative Covariates:** Features such as an elevator, kitchen, guest toilet, and parking space were associated with higher rents. For instance, apartments with upscale furnishing saw a significant impact in Munich in 2019, with rent per sqm increasing by over 16% from normal to upscale furnishing, reflecting the premium placed on such amenities. These findings are consistent with existing literature highlighting the importance of property features and amenities in determining rental values [4].

Also, the study observed that energy efficiency ratings inversely affected rental prices, with higher efficiency ratings correlating with lower rents. This counterintuitive finding suggests that tenants may not fully value energy efficiency in their rental decisions, a phenomenon also noted in previous research [4]. Furthermore, the scatter plot of the continuous variable (energy condition) versus the rent in both the raw and predicted models, as shown in **Figure 2**, **Table 6** and **Table 12**, demonstrates a consistent decreasing trend in both the raw data and model predictions, supporting the validity of this result. This behavior may reflect market dynamics where energy-efficient features are undervalued or not effectively communicated during the rental process.

### 5.3. Contribution of the Study

The contribution of this study can be summarized in the following themes:

- **Advancing Statistical and Mathematical Applications in Real-World Contexts:** This study contributes to applied statistics by demonstrating multiple linear regression and backward stepwise selection to model rent per square meter in Munich. It shows how rigorous statistical techniques can extract insights from real estate data, supporting predictive modeling in housing markets. The research also provides a practical case study for teaching regression analysis in mathematics and data science courses [19] [20] [30] [31], strengthening the integration of mathematics into applied socioeconomic research.
- **Supporting Educational Leadership through Data-Driven Policy:** With increasing housing costs affecting faculty and students alike, this study offers actionable insights for higher education leaders. Identifying key rent-influencing factors (such as energy efficiency, furnishing level, and apartment size) enables university administrators and planners to advocate for informed housing policies. The work underscores the leadership principle of evidence-based decision-making, a cornerstone in educational leadership programs [8].
- **Enhancing Institutional Housing Strategies:** The study's findings directly affect student and faculty housing strategies in Germany, the United States, and other countries. It provides a model that can be replicated in other high-demand university cities facing affordability challenges. In particular, institutions can use the insights to determine which apartment features drive prices and how these impact different stakeholder groups. This evidence-based approach supports targeted interventions in housing negotiations, construction planning, and subsidy designs [6] [9].

- **Cross-National Relevance to Housing and Urban Studies:** Focusing on Munich, a city with housing dynamics comparable to urban centers in the U.S., the research offers a foundation for comparative studies between European and American academic housing environments. It bridges the contextual divide between Germany's state-supported housing initiatives and the market-driven models prevalent in U.S. academia, helping researchers and policymakers draw transnational lessons about affordability, space optimization, and energy use in university housing [5].
- **Bridging Educational Research, Sustainability, and Equity:** This study also supports broader goals in educational leadership by addressing themes of sustainability and equity. Variables such as energy efficiency and modernization year provide insight into how eco-conscious housing design intersects with rent prices. It contributes to ongoing efforts to promote environmentally sustainable housing options for academic communities [7] while addressing disparities in student access to affordable, high-quality housing.

## 6. Conclusion and Implications

This study examined the dynamics of rental apartment prices per square meter in Munich using a robust statistical framework grounded in multiple linear regression with nonlinear covariates. Drawing from an extensive dataset provided by FDZ Ruhr in collaboration with ImmobilienScout24, the research analyzed over 29,000 rental listings across two critical periods, 2015 and 2019. Our findings revealed influential factors influencing rental price variations, including apartment size, furnishing quality, energy efficiency ratings, and availability of amenities such as elevators and parking spaces.

The analysis identified a consistent inverse relationship between apartment size and rent per square meter, affirming that larger apartments tend to command lower per-unit rents. Upscale furnishings and energy-efficient features were strongly associated with higher rental values, emphasizing the market's shift toward sustainable and modern living preferences. Applying log transformation and polynomial terms improved the model's performance and revealed nuanced nonlinear patterns across years, supported by adjusted R-squared values above 0.5 for the 2019 model.

This study provides valuable insights for stakeholders in the real estate, educational leadership, and urban planning sectors, particularly as Munich grapples with housing shortages and rising rent inflation. Policymakers can use these results to identify leverage points for regulating rental markets and implementing incentive structures for energy-efficient housing. Moreover, the educational implications of the modeling approach underscore the importance of integrating data science and urban economics in curriculum development for future housing strategists.

Future research could extend this model across multiple German cities or apply time-series forecasting techniques to predict rent trends beyond 2019. Incorporating spatial econometrics and GIS-based analysis could also enhance under-

standing of geographic rental disparities within the city. Also, a comparative study of major U.S. and European university cities could be considered for further research to deepen the knowledge of rental price trends and inform global policy practice.

### 6.1. Recommendations

Based on the findings of this study, the following recommendations are proposed:

- **University Housing Policy:** Higher education institutions should incorporate rent-influencing factors, like furnishing quality and energy efficiency, into campus housing strategies. This can support better decision-making around subsidies, leasing agreements, and student financial aid.
- **Data-Driven Leadership:** Educational leaders should utilize regression-based evidence to inform housing advocacy and planning. Insights from this study can be integrated into strategic plans to ensure equity and affordability in university accommodations.
- **Urban Planning and Sustainability:** Real estate developers and city planners in high-demand university cities should prioritize modernization and sustainable features in housing designs while acknowledging their nuanced effects on rent and tenant preferences.

### 6.2. Limitations

- **Geographical Scope:** The research is limited to Munich, Germany, which may affect the generalizability of findings to other cities with different regulatory frameworks or housing demands.
- **Temporal Coverage and Data Constraints:** Only two years (2015 and 2019) were analyzed. Trends may differ significantly in the context of more recent economic or policy changes, especially post-pandemic. Also, although the dataset was comprehensive, some potential predictors, like tenant demographics, or other economic indicators, were not included.

### Acknowledgements

This paper builds on the Master's thesis of Ugochukwu Onumadu, supervised by Prof. Ph.D. Claudia Czado, at the Technical University of Munich, Germany. It refines the original findings, explicitly focusing on the Munich rental market.

### Conflicts of Interest

I, Ugochukwu Onumadu, the author of this study, declares that there are no conflicts of interest associated with this publication. No financial or non-financial interests, personal relationships, or institutional affiliations have influenced the content, results, or interpretation of this study.

### References

- [1] Moberg, J. (2017) Outlook of the German Housing Market in 2017. Outlook.

- [2] Lutz, E. (2020) The Housing Crisis as a Problem of Intergenerational Justice: The Case of Germany. *Intergenerational Justice Review*, **6**, Article No. 1.
- [3] Malpezzi, S., *et al.* (2003) Hedonic Pricing Models: A Selective and Applied Review. *Housing Economics and Public Policy*, **1**, 67-89.
- [4] Yoshida, T., Murakami, D. and Seya, H. (2022) Spatial Prediction of Apartment Rent Using Regression-Based and Machine Learning-Based Approaches with a Large Dataset. *The Journal of Real Estate Finance and Economics*, **69**, 1-28.  
<https://doi.org/10.1007/s11146-022-09929-6>
- [5] Brookings Institution (2023) How a University-Community Home-Sharing Collective Is Creating a New Model for Affordable Housing in West Philadelphia.
- [6] U.S. Department of Housing and Urban Development (2023) Worst Case Housing Needs: 2023 Report to Congress.
- [7] Pivo, G. (2022) Green Buildings and Rental Premiums: A Meta-Analysis. *Journal of Sustainable Real Estate*, **14**, 1-16.
- [8] Fullan, M. (2020) *Leading in a Culture of Change*. 2nd Edition, John Wiley & Sons.
- [9] German Academic Exchange Service (DAAD) (2024) Internationalisation Only Successful with Sufficient Living Space for Students.
- [10] Fahrmeir, L., Kneib, T., Lang, S. and Marx, B. (2013) Regression Models. In: Fahrmeir, L., Kneib, T., Lang, S. and Marx, B., Eds., *Regression*, Springer, 21-72.  
[https://doi.org/10.1007/978-3-642-34333-9\\_2](https://doi.org/10.1007/978-3-642-34333-9_2)
- [11] Allen, M.P. (2004) *Understanding Regression Analysis*. Springer Science & Business Media.
- [12] Christensen, R. (1996) *Analysis of Variance, Design, and Regression: Applied Statistical Methods*. CRC Press.
- [13] Christensen, R. (2018) *Analysis of Variance, Design, and Regression: Linear Modeling for Unbalanced Data*. Chapman and Hall/CRC.
- [14] Horton, N.J. and Kleinman, K. (2015) *Using R and RStudio for Data Management, Statistical Analysis, and Graphics*. CRC Press.
- [15] Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, **135**, 370-384.  
<https://doi.org/10.2307/2344614>
- [16] Abraham, B. and Ledolter, J. (2006) *Student Solutions Manual for Introduction to Regression Modeling*. Cengage Learning.
- [17] Ricci, L. (2010) Adjusted-Squared Type Measure for Exponential Dispersion Models. *Statistics & Probability Letters*, **80**, 1365-1368.  
<https://doi.org/10.1016/j.spl.2010.04.019>
- [18] McNeil, K.A., Newman, I. and Kelly, F.J. (1996) *Testing Research Hypotheses with the General Linear Model*. SIU Press.
- [19] Seber, G.A. (2015) *The Linear Model and Hypothesis*. Springer.
- [20] Vik, P. (2014) *Regression, ANOVA, and the General Linear Model: A Statistics Primer*. SAGE Publications. <https://doi.org/10.4135/9781071939024>
- [21] Lin, D.Y., Wei, L.J. and Ying, Z. (2002) Model-Checking Techniques Based on Cumulative Residuals. *Biometrics*, **58**, 1-12.  
<https://doi.org/10.1111/j.0006-341x.2002.00001.x>
- [22] Osborne, J.W. and Waters, E. (2002) Four Assumptions of Multiple Regression That Researchers Should Always Test. *Practical Assessment, Research, and Evaluation*, **8**, Article No. 2.

- 
- [23] Lindsey, J.K. (2000) Applying Generalized Linear Models. Springer Science & Business Media.
- [24] Farrar, D.E. and Glauber, R.R. (1967) Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics*, **49**, 92-107. <https://doi.org/10.2307/1937887>
- [25] Neter, J., Wasserman, W. and Kutner, M.H. (1983) Applied Linear Regression Models. Richard D. Irwin.
- [26] Jones Lang LaSalle (JLL) (2024) Housing Market Overview—H2 2024. <https://www.jll.de/en/trends-and-insights/research/housing-market-overview>
- [27] Rusakov, O.V., Laskin, M.B. and Jaksumbaeva, O.I. (2015) Stochastic Pricing Model for the Real Estate Market: Formation of Log-Normal General Population. *Statistics and Economics*, No. 5, 116-127. <https://doi.org/10.21686/2500-3925-2015-5-116-127>
- [28] Laskin, M. and Rusakov, O. (2023) Prediction of Distributions of Unit Prices for Real Estate Properties on the Basis of the Characteristics of PSI-Processes. *Business Informatics*, **17**, 7-24. <https://doi.org/10.17323/2587-814x.2023.4.7.24>
- [29] D'Acci, L.S. (2023) Is Housing Price Distribution across Cities, Scale Invariant? Fractal Distribution of Settlements' House Prices as Signature of Self-Organized Complexity. *Chaos, Solitons & Fractals*, **174**, Article ID: 113766. <https://doi.org/10.1016/j.chaos.2023.113766>
- [30] Czado and Brechmann (2021) Lecture Slides on GLM, Study Material from the Research Group Mathematical Statistics in the Department of Mathematics at the Technical University Munich Deutschland. <https://www.groups.ma.tum.de/statistics/personen/claudia-czado/forschung/lecture-slides/>
- [31] McConnell, J.R., Short, P.C. and Ross, S.M. (2024) Introductory Statistics: A Contextualized Approach. 4th Edition, Linus Publishing.