Online-Appendix zu

„The Economic Upside of Green Real Estate Investments: Analyzing the Impact of Energy Efficiency on Building Valuation in the Residential Sector“

Timo Deller

Technische Universität München

Junior Management Science 7(3) (2022) 802-825

Appendix

Appendix 1: Summary of variables and their usage in models

Variable	Data codification & transformation	Description of variable	Cold Rent Model	Warm Rent Model	Sales Price Model
Dependent variables					
ln(cold_rent)	Quantitative; log-transformed	The natural logarithm of the cold rent of the building. Unit: EUR	x		
ln(warm_rent)	Quantitative; log-transformed	The natural logarithm of the warm rent of the building. Unit: EUR		x	
ln(sales_price)	Quantitative; log transformed	The natural logarithm of the sales price of the building. Unit: EUR			x
Building-specific independent variables					
EPC-level	Dummy	Indicating the German EPC level ranging from A+ to H. The reference value of the dummy is set to D. Overall, 9 levels.	x	x	x
ln(living_space)	Quantitative; log-transformed	The natural logarithm of the living space of the building. Unit: m^2	x	x	x
no_rooms	Dummy	Indicating the number of rooms ranging as a categorical feature of the building.	x	x	x
furnished	Dummy	Whether the building comes with ready to use furniture or not.	x	x	
refurbished	Dummy	Whether the building has been refurbished since the original construction year.	x	x	x
first_occupancy	Dummy	Whether no one has lived in the building before or not.	x	x	x
landmarked_building	Dummy	Whether the building falls under the German “Denkmalschutz” or not.	x	x	x
elevator	Dummy	Whether an elevator is present in the building or not.	x	x	x
parking_space	Dummy	Whether a parking space is available or not.	x	x	x
building_type	Dummy	Controlling for the different building types: e.g. ground floor apartment, penthouse, multi-family.	x	x	x

construction_year	Dummy	Controlling for the different construction years with time periods of 10 years.	x	x	x
rent_status	Dummy	Whether the building is currently let to a tenant or not.			x
Contract-specific independent variable					
commission_free	Dummy	Whether a commission is to be paid to a broker when buying the building or not.			x
Location-specific independent variable					
γ	Dummy	Controlling for the building location on a postal code level. This helps to control for population density, purchasing power and other factors.	x	x	x
Time-specific independent variable					
δ	Dummy	Controlling for the different upload dates on a monthly level starting with 01/2019 and ending with 12/2020.	x	x	x

Appendix 2: Correlation matrix rent data

		i)	ii)	iii)	iv)	v)	vi)	vii)	viii)	ix)	x)	xi)	xii)
i)	Cold Rent	1											
ii)	Warm Rent	0.995	1										
iii)	Energy Consumption	-0.24	-0.23	1									
iv)	Living Space	0.79	0.80	-0.14	1								
v)	Number of Rooms	0.63	0.64	-0.07	0.86	1							
vi)	Furnished	0.01	0.01	0.02	-0.09	-0.10	1						
vii)	Refurbished	-0.03	-0.03	0.21	-0.03	-0.01	0.03	1					
viii)	First Occupancy	0.18	0.17	-0.23	0.06	0.04	-0.02	0.04	1				
ix)	Landmarked Building	-0.00	-0.00	0.01	-0.01	-0.01	-0.00	0.01	-0.01	1			
x)	Elevator	0.20	0.21	-0.33	0.01	-0.08	0.04	-0.11	0.18	0.01	1		
xi)	Parking Space	0.21	0.22	-0.25	0.25	0.19	-0.04	-0.10	0.08	-0.02	0.12	1	
xii)	Year of Construction	0.23	0.24	-0.55	0.17	0.07	-0.01	-0.27	0.22	-0.04	0.35	0.41	1

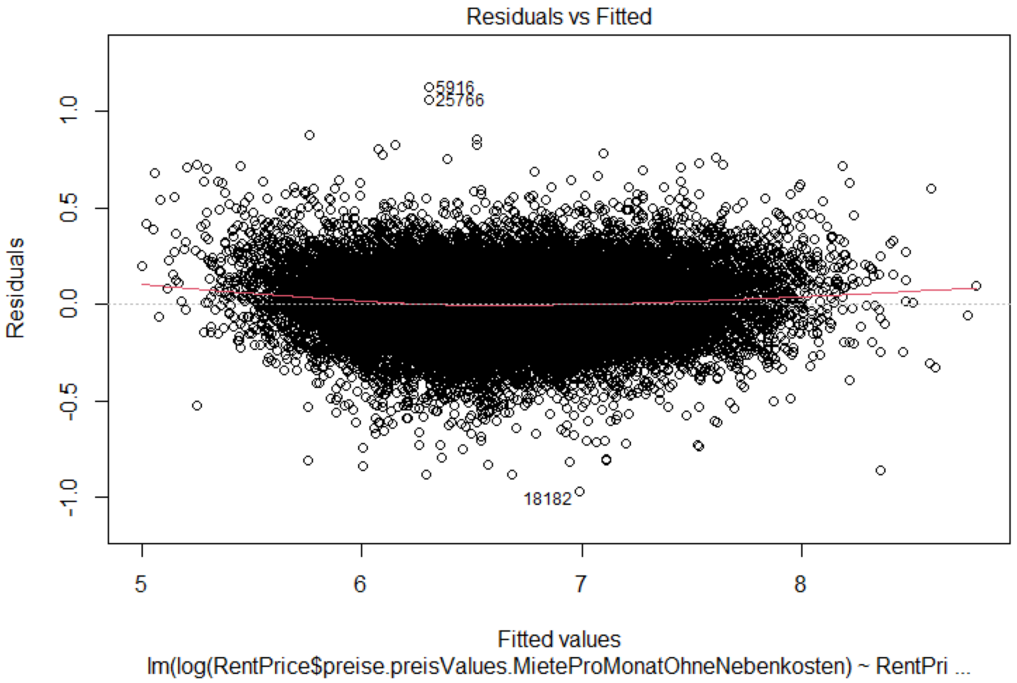
*Pearson's correlation coefficient rounded to two digits behind the comma.

Appendix 3: Correlation matrix sales data

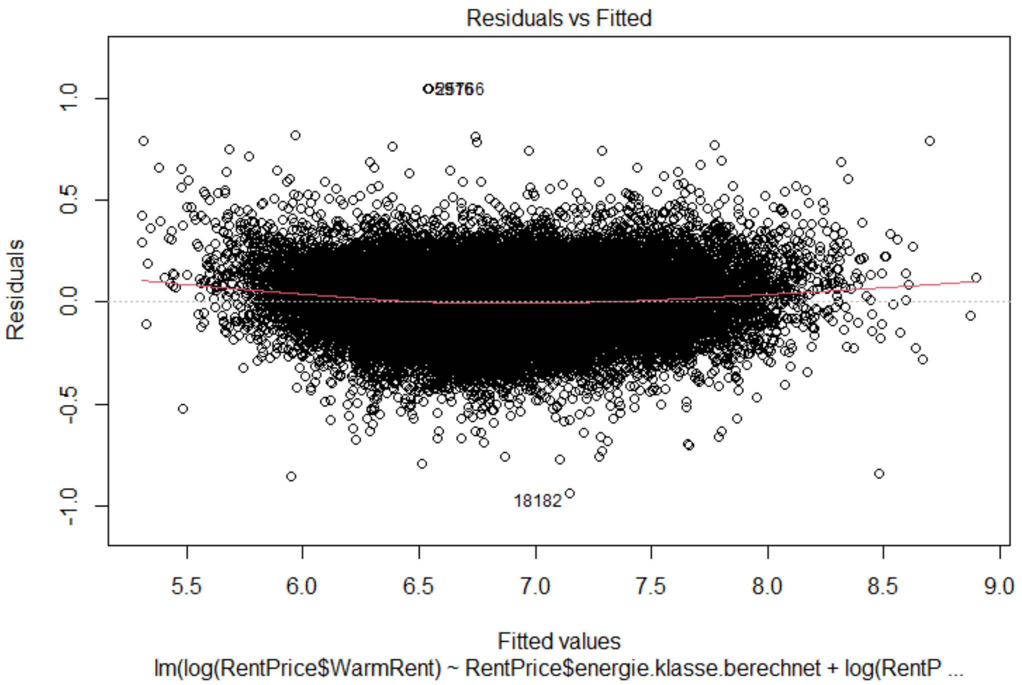
		i)	ii)	iii)	iv)	v)	vi)	vii)	viii)	ix)	x)	xi)	xii)
i)	Sales Price	1											
ii)	Energy Consumption	-0.10	1										
iii)	Living Space	0.67	0.06	1									
iv)	Number of Rooms	0.55	0.14	0.90	1								
v)	Refurbished	-0.01	0.07	-0.03	-0.02	1							
vi)	First Occupancy	0.06	-0.35	-0.07	-0.10	-0.09	1						
vii)	Landmarked Building	0.02	0.01	-0.00	-0.00	0.06	-0.00	1					
viii)	Elevator	-0.04	-0.23	-0.23	-0.26	0.01	0.22	0.02	1				
ix)	Parking Space	0.05	-0.04	0.03	0.02	0.03	0.10	-0.02	0.00	1			
x)	Year of Construction	0.07	-0.67	-0.08	-0.16	-0.17	0.38	-0.08	0.20	0.13	1		
xi)	Active Lease	-0.06	0.04	-0.00	0.02	0.05	-0.11	0.04	0.07	-0.03	-0.08	1	
xii)	No Commission	0.02	-0.21	-0.04	-0.05	0.00	0.34	0.05	0.07	0.07	0.21	-0.05	1

*Pearson's correlation coefficient rounded to two digits behind the comma.

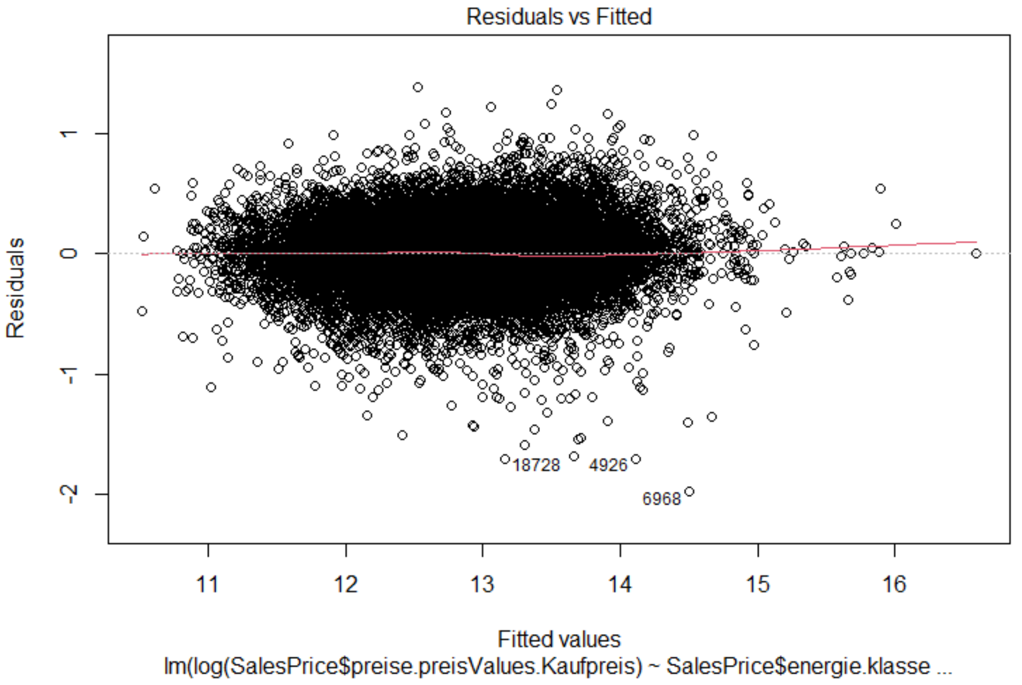
Appendix 4: Residual Plot Cold Rent Model



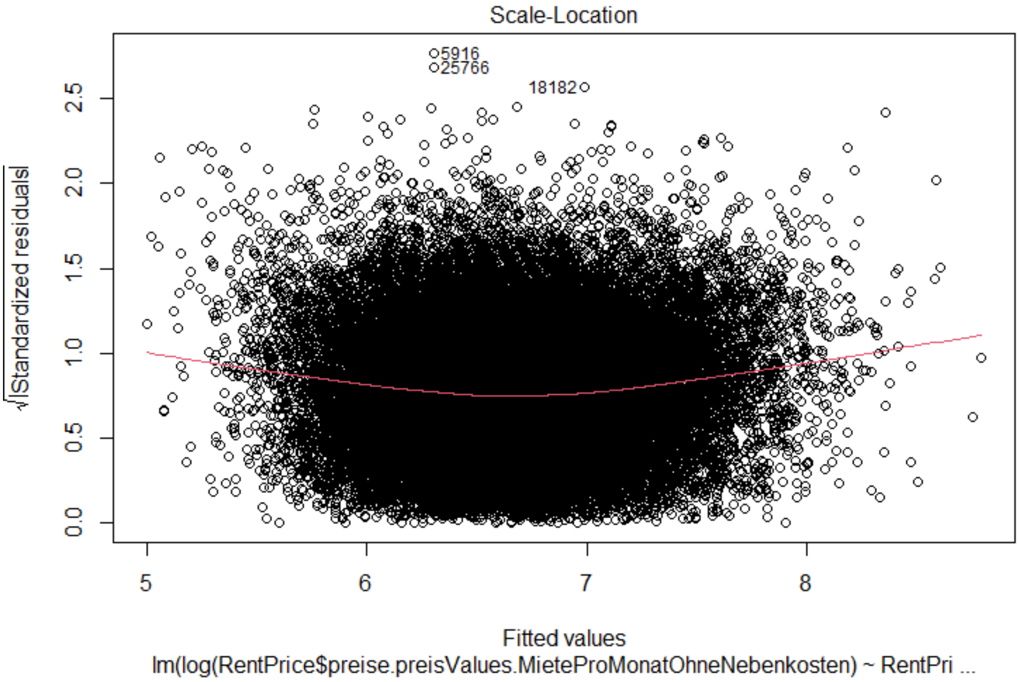
Appendix 5: Residual Plot Warm Rent Model



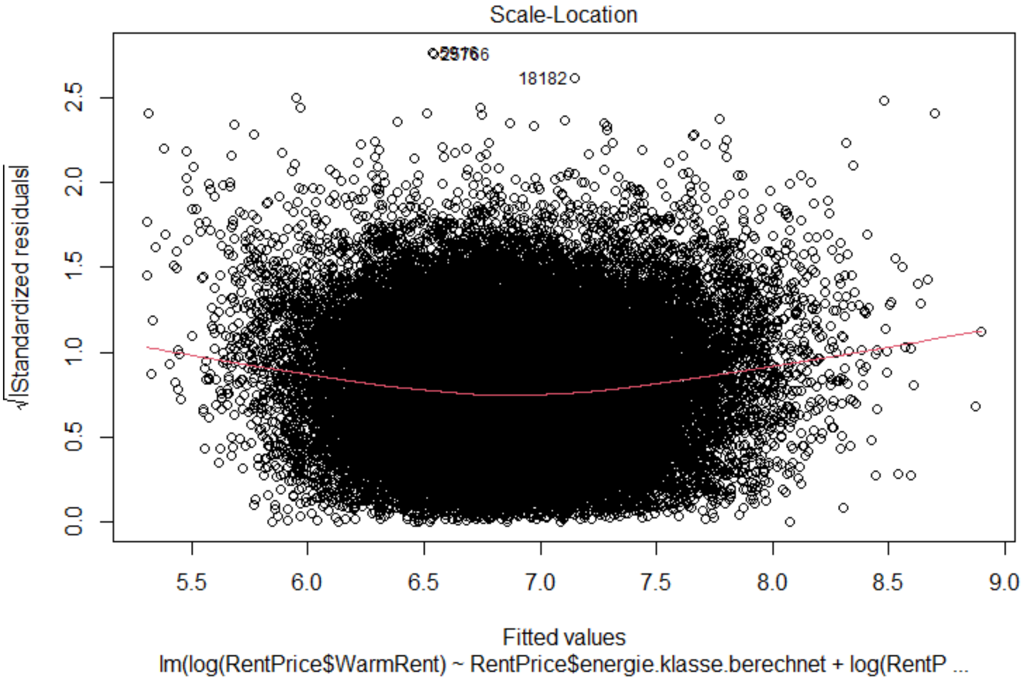
Appendix 6: Residual Plot Sales Price Model



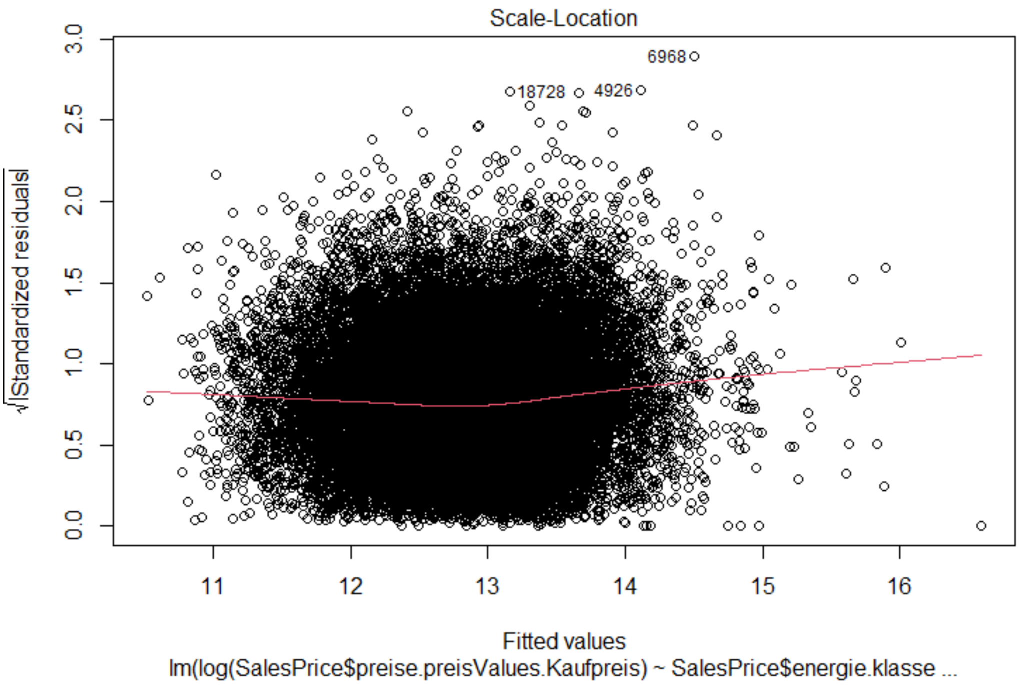
Appendix 7: Scale-Location Plot Cold Rent Model



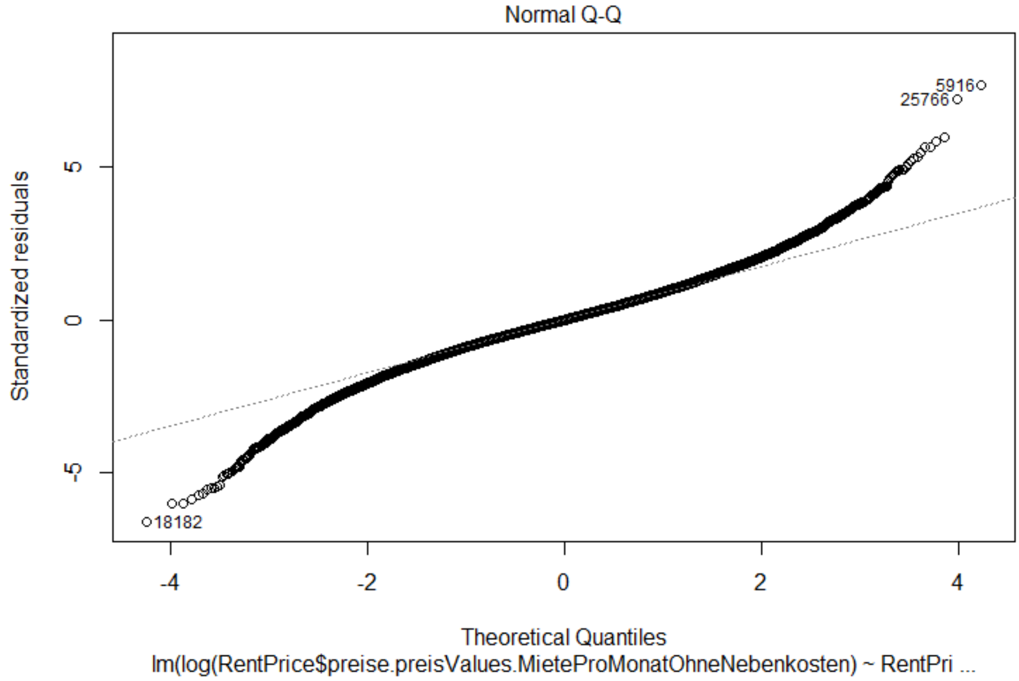
Appendix 8: Scale-Location Plot Warm Rent Model



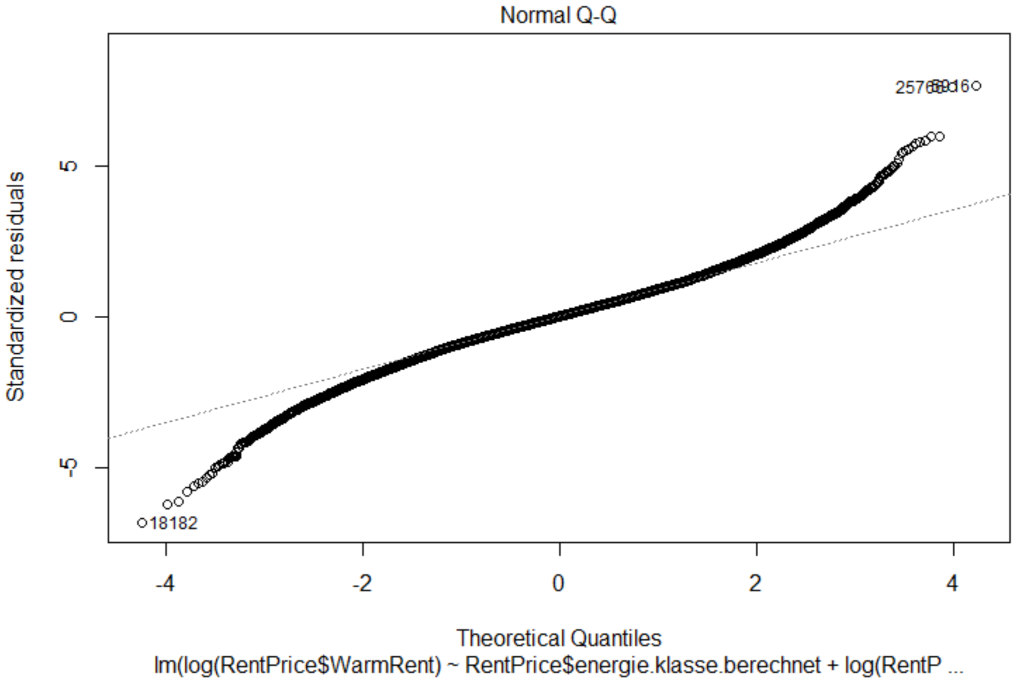
Appendix 9: Scale-Location Plot Sales Price Model



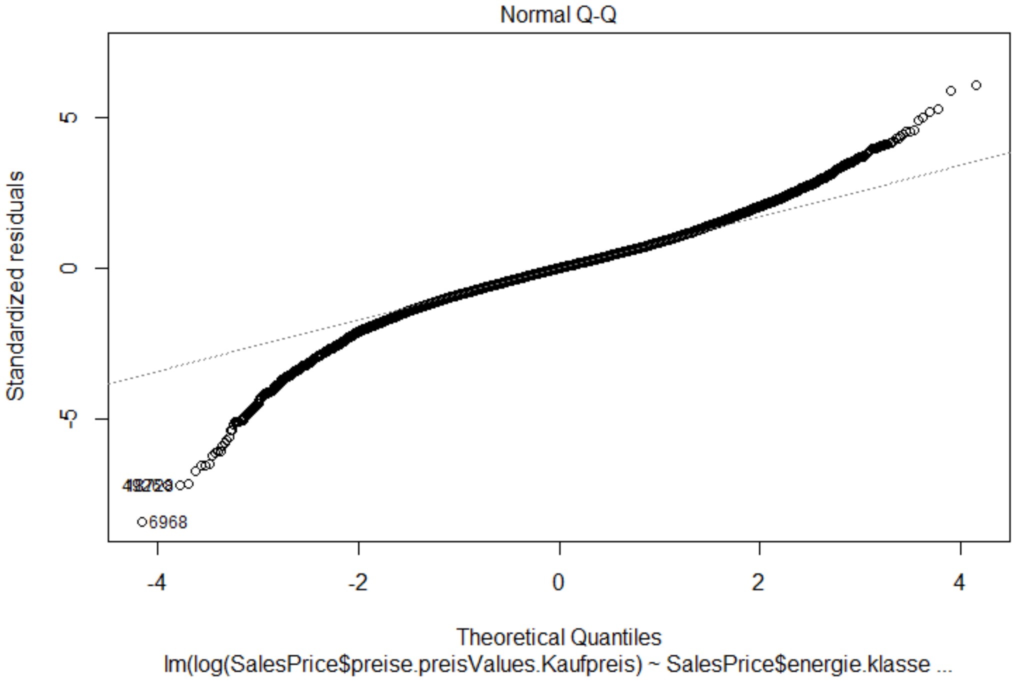
Appendix 10: Normal Quantile-Quantile Plot Cold Rent Model



Appendix 11: Normal Quantile-Quantile Plot Warm Rent Model



Appendix 12: Normal Quantile-Quantile Plot Sales Price Model



Appendix 13: GVIF Calculation Rent Price Data

Variable	GVIF	DF	GVIF^(1/(2*DF))
Energy Consumption	4.317794	8	1.095731
Living Space	5.415959	1	2.327221
Number of Rooms	12.031062	22	1.058162
Furnished	1.052056	1	1.025698
Refurbished	1.190888	1	1.091278
First Occupancy	1.315426	1	1.146920
Landmarked Building	1.041974	1	1.020771
Elevator	1.640310	1	1.280746
Parking Space	1.408612	1	1.186850
Construction Year	11.817925	15	1.085804
Object Type	3.127981	9	1.065405
Date Upload	1.486965	23	1.008662
Postal Code	13.188097	455	1.002838

Appendix 14: GVIF Calculation Sales Price Data

Variable	GVIF	DF	GVIF^(1/(2*DF))
Energy Consumption	9.108901	8	1.148065
Living Space	7.813896	1	2.795335
Number of Rooms	92.776600	63	1.036608
Refurbished	1.114746	1	1.055815
First Occupancy	1.917547	1	1.384755
Landmarked Building	1.079626	1	1.039051
Elevator	1.746292	1	1.321473
Parking Space	1.152916	1	1.073739
Rent Status	1.186996	1	1.089494
Commission Free	1.236495	1	1.111978
Construction Year	15.834914	15	1.085804
Object Type	11.148133	9	1.065405
Date Upload	1.944487	23	1.008662
Postal Code	64.307732	479	1.002838

Appendix 15: R – Script

```
## Version 1.0 - Written by Timo Deller ##
## R - Script for Empirical Analysis
## Bachelor Thesis ##
## Rent Prices ##

## If no package is given, then the function belongs to the {base} or {stats} package in R.
## All linear models were computed using the lm() function of the {stats} package.

## Rent price data import

# Load Data into R
RentPriceOriginal <- read.csv(file.choose(), header = T)
# Duplicate original data set.
RentPrice <- RentPriceOriginal
# Choose relevant variables for the Cold/Warm Rent Price Model
RentPrice <- RentPrice[,c(7, 22, 23, 32, 38, 53, 66, 69, 76, 78, 83, 89, 103, 109, 115)]
#N = 244 277

## Plausibility check of rent price data ##
## Deletion of observations with information entered that is not plausible or incomplete and thus cannot be used for the analysis.##

## Construction year plausibility check:

# Deletion of observations with a construction year that is greater than 2021 or when no construction year given.
# Real Estate development firms use internet platforms for advertisement and enter a future point in time as construction year.
# This condition gets rid of these advertisements. The year 2021 is included since it is plausible that buildings being finished in 2021 are already sold/let on the market in the year of 2020, the time period considered in the analysis.
RentPrice <- subset(RentPrice, RentPrice$jahreszahlen.BAUJAHR < 2022)
```

```
#N = 124 670

# Deletion of observations with construction year smaller than 1871. The year 1871 marks the beginning of the Gründerzeit in Germany.
# This was accompanied by a wide introduction of industrialization and change in construction. Further, excluding old buildings helps “To circumvent the problem of unobserved refurbishment of the historical building stock, (…)”[160]
RentPrice <- subset(RentPrice, RentPrice$jahreszahlen.BAUJAHR > 1870)
#N = 123 279

## Energy consumption plausibility check:

# Deletion of all observations that do not have a numeric value for their energy consumption included.
RentPrice <- RentPrice[!is.na(RentPrice$energieausweis.energieKennwerte.ENERGIEKENNWERT.kennwert),]
#N = 64 495

# Deletion of all observations that put 5 or less as energy consumption. This is almost impossible to achieve and suggests wrong information entered in the offer.[161]
RentPrice <- RentPrice[(RentPrice$energieausweis.energieKennwerte.ENERGIEKENNWERT.kennwert > 5),]
#N = 63 986

# Deletion of all observations that have an energy consumption of 500 kwh/m^2 per annum or more entered in the offer.
# An energy consumption equal to or above of this value does not seem plausible.[162]
# It either suggests the value was entered as a place holder, is incorrect or the building has substantial structural deficits and should not be considered in the analysis.
```

[160] Cajias / Fuerst / Bienert (2019), p. 184.
[161] Cf. Green-Towers Sustainable High-Rises GmbH (2021).
[162] Cf. Green-Towers Sustainable High-Rises GmbH (2021).

```
RentPrice <- RentPrice[(RentPrice$energieausweis.energieKennwerte.ENERGIEKENNWERT.kennwert < 500),]
#N = 63 901

## Living space per room plausibility check:

# Deletion of all observations with an unplausible ratio of living space to number of rooms. Unplausible is defined as 100 m^2 per room or more. [163]
RentPrice$mperroom <- RentPrice$flÃ.chen.flÃ.cheValues.Wohnflaeche/RentPrice$anzahlen.ZIMMER
RentPrice <- RentPrice[(RentPrice$mperroom < 100),]
#N = 63 878

# Deletion of observations with 8 m^2 or less of living space per room on average. As bathrooms and the kitchen are not included in the number of rooms, anything below this value does not seem plausible and would be difficult to construct.[164]
RentPrice <- RentPrice[(RentPrice$mperroom > 8),]
#N = 63 860

## Rent price per m^2 plausibility check:

# Deletion of observations with costs equal to or below 4 EUR per m^2 of living space.
# According to Zeit, the lowest rent in all of Germany was 3.93 EUR at the end of 2019.[165]
# As this data set is centered around the Rhein-Main Region, values equal to the lowest rents in Germany do not seem plausible.
RentPrice$costperm <- RentPrice$preise.preisValues.MieteProMonatOhneNebenkosten/RentPrice$flÃ.chen.flÃ.che
Values.Wohnflaeche
RentPrice <- RentPrice[(RentPrice$costperm > 4),]
#N = 63 810
```

[163] Cf. Green-Towers Sustainable High-Rises GmbH (2021).

[164] Cf. Green-Towers Sustainable High-Rises GmbH (2021).

[165] Cf. Blickle et al. (2019).

```
# Due to the possibility of fully furnished houses, introducing an upper limit is not given.

## Operating costs to cold rent ration plausibility check:

# Deletion of observations with a ratio greater or equal to 0.5 and smaller or equal to 0.1. [166]
# A ratio above 0.5 does not seem plausible, even for energy inefficient homes.
# A ratio that is 0.1 or smaller does not seem plausible. This would mean that less than 10% is
added to the cold rent. Such observations most likely have the warm rent already entered in the
cold rent variable, as a lot of them state zero operational costs. With such information entered,
they cannot be used for the different models to answer the stated hypotheses.
RentPrice$coldtowarm                                                                    <-
RentPrice$preise.preisValues.NebenkostenProMonat/RentPrice$preise.preisValues.MietePro
MonatOhneNebenkosten
RentPrice <- RentPrice[(RentPrice$coldtowarm > 0.1),]
RentPrice <- RentPrice[(RentPrice$coldtowarm < 0.5),]
# N = 59 866

## Codification of independent variables

# Add control variable for the construction year.
RentPrice$baujahr.abschnitte <- cut(RentPrice$jahreszahlen.BAUJAHR, c(2021,
                                           2011,
                                           2001,
                                           1991,
                                           1981,
                                           1971,
                                           1961,
                                           1951,
                                           1941,
                                           1931,
                                           1921,
                                           1911,
                                           1901,
```

[166] Cf. Green-Towers Sustainable High-Rises GmbH (2021).

```
                                        1891,
                                        1881,
                                        1871,
                                        1861))

# The dummy variables are changed to numeric so that they can be used for the calculation of
descriptive statistics.

# Dummy variable: refurbished
RentPrice$saniert[!RentPrice$saniert =="true"] <- 0
RentPrice$saniert[RentPrice$saniert =="true"] <- 1
RentPrice$saniert <- as.numeric(RentPrice$saniert)

# Dummy variable: first occupancy
RentPrice$erstbezug[!RentPrice$erstbezug =="true"] <- 0
RentPrice$erstbezug[RentPrice$erstbezug =="true"] <- 1
RentPrice$erstbezug <- as.numeric(RentPrice$erstbezug)

# Dummy variable: landmarked building
RentPrice$denkmalschutz[!RentPrice$denkmalschutz =="true"] <- 0
RentPrice$denkmalschutz[RentPrice$denkmalschutz =="true"] <- 1
RentPrice$denkmalschutz <- as.numeric(RentPrice$denkmalschutz)

# Dummy variable: elevator
RentPrice$aufzug[!RentPrice$aufzug ==""] <- 1
RentPrice$aufzug[RentPrice$aufzug ==""] <- 0
RentPrice$aufzug <- as.numeric(RentPrice$aufzug)

# Dummy variable: furnished
RentPrice$mÃ.bliert[!RentPrice$mÃ.bliert ==""] <- 1
RentPrice$mÃ.bliert[RentPrice$mÃ.bliert ==""] <- 0
RentPrice$mÃ.bliert <- as.numeric(RentPrice$mÃ.bliert)

# Dummy variable: parking space
```

```
RentPrice$stellplatz[!RentPrice$stellplatz ==""] <- 1
RentPrice$stellplatz[RentPrice$stellplatz ==""] <- 0
RentPrice$stellplatz <- as.numeric(RentPrice$stellplatz)

# Dummy variable: building type
# Deletion of observations that do not have a building type entered or "other" entered as building type.
# As it is unclear what "other" stands for, no clear statement can be made about those observations. Further "other" might also include building types already controlled for and thus impact the effectiveness of the control variable.
RentPrice <- RentPrice[!(RentPrice$objektTyp == ""),]
RentPrice <- RentPrice[!(RentPrice$objektTyp == "Sonstige"),]
RentPrice$objektTyp <- as.factor(RentPrice$objektTyp)
# Changing reference value to "apartment".
RentPrice$objektTyp <- relevel(RentPrice$objektTyp, ref = "Etagenwohnung")
#N = 47 818

# Calculation of the EPC dummy variable for each observation according to the current law, the GEG (01/02/2021).
RentPrice$energie.klasse.berechnet <- RentPrice$energieausweis.energieKennwerte.ENERGIEKENNWERT.kennwert
RentPrice$energie.klasse.berechnet <- cut(RentPrice$energie.klasse.berechnet, breaks = c(0,30,50,75,100,130,160,200,250,500), labels = c("A+","A","B","C","D","E","F","G","H"))

# Control variable for the upload date of the observation (monthly control).
RentPrice$dateUpload <- RentPrice$firstSeenDay
RentPrice$dateUpload <- as.Date.character(RentPrice$dateUpload)
RentPrice$dateUpload <- format(as.Date(RentPrice$dateUpload), "%Y-%m")
#Delete all observations that were uploaded before January 2019.
RentPrice <- RentPrice[(RentPrice$dateUpload>"2018-12"),]
#N = 45 528

# There are no observations with an upload date later than 12/2020 present in the data set.
```

```
#Calculation of the warm rent
RentPrice$WarmRent<-  RentPrice$preise.preisValues.MieteProMonatOhneNebenkosten  +
RentPrice$preise.preisValues.NebenkostenProMonat

# Deletion of any left-over incomplete cases
RentPrice <- RentPrice[complete.cases(RentPrice[ , 1:22]),]
#N = 44 442

## Plots and descriptive statistics

# Plot EPC levels
# Not the plot used in the final thesis document. Just used as an overview.
plot(RentPrice$energie.klasse.berechnet)
# Data used for the bar plot shown in Figure 2 in final thesis document.
summary(RentPrice$energie.klasse.berechnet)

# Descriptive Statistics
summary(RentPrice)
print(summary(RentPrice$preise.preisValues.MieteProMonatOhneNebenkosten), digits=20)
sd(RentPrice$preise.preisValues.MieteProMonatOhneNebenkosten)
print(summary(RentPrice$WarmRent), digits=20)
sd(RentPrice$WarmRent)
sd(RentPrice$energieausweis.energieKennwerte.ENERGIEKENNWERT.kennwert)
sd(RentPrice$flÃ.chen.flÃ.cheValues.Wohnflaeche)
sd(RentPrice$anzahlen.ZIMMER)
sd(RentPrice$mÃ.bliert)
sd(RentPrice$saniert)
sd(RentPrice$erstbezug)
sd(RentPrice$denkmalschutz)
sd(RentPrice$aufzug)
sd(RentPrice$stellplatz)
sd(RentPrice$jahreszahlen.BAUJAHR)

# Calculation of the correlation matrix.
```

```
RentPriceCor <- RentPrice[,c(13,22,5, 7,2,9,10,6,4,3,11,8)]
RentPriceCorMatrix <- cor(RentPriceCor)
# Export as CSV file.
write.csv(RentPriceCorMatrix, file = "RentPriceCorMatrix.csv")

## Rent price linear models

# Change class of variable to factor
RentPrice$objektTyp <- as.factor(RentPrice$objektTyp)
RentPrice$aufzug <- as.factor(RentPrice$aufzug)
RentPrice$denkmalschutz <- as.factor(RentPrice$denkmalschutz)
RentPrice$erstbezug <- as.factor(RentPrice$erstbezug)
RentPrice$saniert <- as.factor(RentPrice$saniert)
RentPrice$stellplatz <- as.factor(RentPrice$stellplatz)
RentPrice$mÃ.bliert <- as.factor(RentPrice$mÃ.bliert)
RentPrice$addr.plz <- as.factor(RentPrice$addr.plz)
RentPrice$dateUpload <- as.factor(RentPrice$dateUpload)
RentPrice$anzahlen.ZIMMER <- as.factor(RentPrice$anzahlen.ZIMMER)

# Change the reference value of the EPC level to D
RentPrice$energie.klasse.berechnet <- relevel(RentPrice$energie.klasse.berechnet, ref = "D")

# Cold Rent Model
# Definition of the multiple linear regression model
LinearModelColdRent <- lm(formula =
log(RentPrice$preise.preisValues.MieteProMonatOhneNebenkosten) ~
            RentPrice$energie.klasse.berechnet +
            log(RentPrice$flÃ.chen.flÃ.cheValues.Wohnflaeche) +
            RentPrice$anzahlen.ZIMMER +
            RentPrice$baujahr.abschnitte +
            RentPrice$mÃ.bliert +
            RentPrice$saniert +
            RentPrice$erstbezug +
            RentPrice$denkmalschutz +
                    RentPrice$aufzug +
                    RentPrice$stellplatz +
                    RentPrice$objektTyp +
                    RentPrice$dateUpload +
                    RentPrice$addr.plz)

# Summary of the Cold Rent Model
summary(LinearModelColdRent)

# Plots of Cold Rent Model to check assumptions
# Package:{graphics}
plot(LinearModelColdRent)

# Calculation of robust standard errors
# Using coeftest() from the package {lmtest} and the heteroscedasticity-consistent covariance matrix estimation vcovHC() from the package {sandwich}. The estimator used is "HC0" that gives White's estimator. (Package {zoo} needed for package {lmtest}.)
library(lmtest)
library(sandwich)
coeftest(LinearModelColdRent , vcov = vcovHC(LinearModelColdRent , type = "HC0"))

# Exogeneity-Assumption: The mean of the error term is equal to zero.
mean(LinearModelColdRent$residuals)

# Autocorrelation-Assumption: There exists no covariance between error terms.
# Using durbinWatsonTest() from the {car} package for computation.
library(car)
durbinWatsonTest(LinearModelColdRent)

# Multicollinearity-Assumption: There exists no perfect multicollinearity between explanatory variables.
# Using vif() from the {car} package for computation.
vif(LinearModelColdRent)
```

```
# Warm Rent Model
# Definition of the multiple linear regression model
LinearModelWarmRent <- lm(formula = log(RentPrice$WarmRent) ~
                        RentPrice$energie.klasse.berechnet +
                        log(RentPrice$flÃ.chen.flÃ.cheValues.Wohnflaeche) +
                        RentPrice$anzahlen.ZIMMER +
                        RentPrice$baujahr.abschnitte +
                        RentPrice$mÃ.bliert +
                        RentPrice$saniert +
                        RentPrice$erstbezug +
                        RentPrice$denkmalschutz +
                        RentPrice$aufzug +
                        RentPrice$stellplatz +
                        RentPrice$objektTyp +
                        RentPrice$dateUpload +
                        RentPrice$addr.plz)

# Summary of the Cold Rent Model
summary(LinearModelWarmRent)

# Plots of Cold Rent Model to check assumptions
# Package:{graphics}
plot(LinearModelWarmRent)

# Calculation of robust standard errors
# Using coeftest() from the package {lmtest} and the heteroscedasticity-consistent covariance
matrix estimation vcovHC() from the package {sandwich}. The estimator used is "HC0" that
gives White's estimator. (Package {zoo} needed for package {lmtest}.)
coeftest(LinearModelWarmRent , vcov = vcovHC(LinearModelWarmRent , type = "HC0"))

# Exogeneity-Assumption: The mean of the error term is equal to zero.
mean(LinearModelWarmRent$residuals)

# Autocorrelation-Assumption: There exists no covariance between error terms.
# Using durbinWatsonTest() from the {car} package for computation.
durbinWatsonTest(LinearModelWarmRent)

# Multicollinearity-Assumption: There exists no perfect multicollinearity between explanatory variables.
# Using vif() from the {car} package for computation.
# Will result in the same results as vif(LinearModelColdRent) because they are based on the same data set.
# Thus, only one of the computations must be done. Added here for completion and for the case that only the Warm Rent Model is looked at.
vif(LinearModelWarmRent)

## END OF R - SCRIPT FOR RENT PRICE ANALYSIS ##



## Version 1.0 - Written by Timo Deller ##
## R - Script for Empirical Analysis
## Bachelor Thesis ##
## Sales Prices ##

## If no package is given, then the function belongs to the {base} or {stats} package in R.
## All linear models were computed using the lm() function of the {stats} package.

## Sales price data import

# Load Data into R.
SalesPriceOriginal <- read.csv(file.choose(), header = T)
# Duplicate original data set.
SalesPrice <- SalesPriceOriginal
# Choose relevant variables for the Sales Price Model
SalesPrice <- SalesPrice[,c(7, 22, 23, 32, 38, 53, 66, 69, 78, 83, 85, 89, 98, 113, 115)]
# N = 123 308

## Plausibility check of sales price data ##
## Deletion of observations with information entered that is not plausible or incomplete and thus cannot be used for the analysis.##

## Construction year plausibility check:

# Deletion of observations with a construction year that is greater than 2021 or when no construction year given.
# Real Estate development firms use internet platforms for advertisement and enter a future point in time as construction year.
# This condition gets rid of these advertisements. The year 2021 is included since it is plausible that buildings being finished in
# 2021 are already sold on the market in the year of 2020, the time period considered in the analysis.
SalesPrice <- subset(SalesPrice, SalesPrice$jahreszahlen.BAUJAHR < 2022)
# N = 98 104

# Deletion of observations with construction year smaller than 1871. The year 1871 marks the beginning of the Gründerzeit in Germany.
# This was accompanied by a wide introduction of industrialization and change in construction. Further, excluding old buildings helps “To circumvent the problem of unobserved refurbishment of the historical building stock, (…)”[167]
SalesPrice <- subset(SalesPrice, SalesPrice$jahreszahlen.BAUJAHR > 1870)
# N = 96 451

## Energy consumption plausibility check:

# Deletion of all observations that do not have a numeric value for their energy consumption included.
SalesPrice <- SalesPrice[!is.na(SalesPrice$energieausweis.energieKennwerte.ENERGIEKENNWERT.kennwert),]
# N = 44 979
```

[167] Cajias, Fuerst and Bienert (2019), p. 184.

```
# Deletion of all observations that put 5 or less as energy consumption. This is almost impossible to achieve and suggests wrong information entered in the offer. [168]
SalesPrice <- SalesPrice[(SalesPrice$energieausweis.energieKennwerte.ENERGIEKENNWERT.kennwert > 5),]
# N = 44 582

# Deletion of all observations that have an energy consumption of 500 kwh/m^2 per annum or more entered in the offer.
# An energy consumption equal to or above of this value does not seem plausible.[169]
# It either suggests the value was entered as a place holder, is incorrect or the building has substantial structural deficits and should not be considered in the analysis.
SalesPrice <- SalesPrice[(SalesPrice$energieausweis.energieKennwerte.ENERGIEKENNWERT.kennwert < 500),]
# N = 44 377

## Living space per room plausibility check:

# Deletion of all observations with an unplausible ratio of living space to number of rooms. Unplausible is defined as 100 m^2 per room or more.[170]
SalesPrice$mperroom <- SalesPrice$flÃ.chen.flÃ.cheValues.Wohnflaeche/SalesPrice$anzahlen.ZIMMER
SalesPrice <- SalesPrice[(SalesPrice$mperroom< 100),]
# N = 44 358

# Deletion of observations with 8 m^2 or less of living space per room on average. As bathrooms and the kitchen are not included in the number of rooms, anything below this value does not seem plausible and would be difficult to construct.[171]
SalesPrice <- SalesPrice[(SalesPrice$mperroom > 8),]
# N = 44 351
```

[168] Cf. Green-Towers Sustainable High-Rises GmbH (2021).
[169] Cf. Green-Towers Sustainable High-Rises GmbH (2021).
[170] Cf. Green-Towers Sustainable High-Rises GmbH (2021).
[171] Cf. Green-Towers Sustainable High-Rises GmbH (2021).

```
## Sales price per m^2 plausibility check:

# Deletion of all observations with a sales price per m^2 that is equal to or below 500 EUR.
# A sales price of that value does not seem plausible. In 2017 the lowest prices per m^2 in Germany were around 670 EUR according to Statista.[172]
# Taking this into account and applying a conservative approach, it is reasonable to delete observations with a price of 500 EUR per m^2 or below.
SalesPrice$costperm <- SalesPrice$preise.preisValues.Kaufpreis/SalesPrice$flÃ.chen.flÃ.cheValues.Wohnflaeche
SalesPrice <- SalesPrice[(SalesPrice$costperm > 500),]
# N = 44 246

# The observations around the maximum sales price per m^2 do not show unplausible values.

## Codification of independent variables

# Add control variable for the construction year.
SalesPrice$baujahr.abschnitte <- cut(SalesPrice$jahreszahlen.BAUJAHR,
                                     c(2021,
                                       2011,
                                       2001,
                                       1991,
                                       1981,
                                       1971,
                                       1961,
                                       1951,
                                       1941,
                                       1931,
                                       1921,
                                       1911,
                                       1901,
                                       1891,
```

[172] Cf. Statista Research Department (2020)

```
                1881,
                1871,
                1861))

# The dummy variables are changed to numeric so that they can be used for the calculation of descriptive statistics.

# Dummy variable: refurbished
SalesPrice$saniert[!SalesPrice$saniert =="true"] <- 0
SalesPrice$saniert[SalesPrice$saniert =="true"] <- 1
SalesPrice$saniert <- as.numeric(SalesPrice$saniert)

# Dummy variable: first occupancy
SalesPrice$erstbezug[!SalesPrice$erstbezug =="true"] <- 0
SalesPrice$erstbezug[SalesPrice$erstbezug =="true"] <- 1
SalesPrice$erstbezug <- as.numeric(SalesPrice$erstbezug)

# Dummy variable: landmarked building
SalesPrice$denkmalschutz[!SalesPrice$denkmalschutz =="true"] <- 0
SalesPrice$denkmalschutz[SalesPrice$denkmalschutz =="true"] <- 1
SalesPrice$denkmalschutz <- as.numeric(SalesPrice$denkmalschutz)

# Dummy variable: elevator
SalesPrice$aufzug[!SalesPrice$aufzug ==""] <- 1
SalesPrice$aufzug[SalesPrice$aufzug ==""] <- 0
SalesPrice$aufzug <- as.numeric(SalesPrice$aufzug)

# Dummy variable: commission free
SalesPrice$preise.provisionsfrei[!SalesPrice$preise.provisionsfrei =="true"] <- 0
SalesPrice$preise.provisionsfrei[SalesPrice$preise.provisionsfrei =="true"] <- 1
SalesPrice$preise.provisionsfrei <- as.numeric(SalesPrice$preise.provisionsfrei)

# Dummy variable: rent status
SalesPrice$vermietet[!SalesPrice$vermietet =="true"] <- 0
```

```
SalesPrice$vermietet[SalesPrice$vermietet =="true"] <- 1
SalesPrice$vermietet <- as.numeric(SalesPrice$vermietet)

# Dummy variable: parking space
SalesPrice$stellplatz[!SalesPrice$stellplatz ==""] <- 1
SalesPrice$stellplatz[SalesPrice$stellplatz ==""] <- 0
SalesPrice$stellplatz <- as.numeric(SalesPrice$stellplatz)

# Dummy variable: building type
# Deletion of observations that do not have a building type entered or "other" entered as building type.
# As it is unclear what "other" stands for, no clear statement can be made about those observations. Further "other" might also include building types already controlled for and thus impact the effectiveness of the control variable.
SalesPrice <- SalesPrice[!(SalesPrice$objektTyp == ""),]
SalesPrice <- SalesPrice[!(SalesPrice$objektTyp == "Sonstige"),]
SalesPrice$objektTyp <- as.factor(SalesPrice$objektTyp)
# Changing reference value to "apartment".
SalesPrice$objektTyp <- relevel(SalesPrice$objektTyp, ref = "Etagenwohnung")
# N = 36 826

# Calculation of the EPC dummy variable for each observation according to the current law, the GEG (01/02/2021).
SalesPrice$energie.klasse.berechnet <- SalesPrice$energieausweis.energieKennwerte.ENERGIEKENNWERT.kennwert
SalesPrice$energie.klasse.berechnet <- cut(SalesPrice$energie.klasse.berechnet, breaks = c(0,30,50,75,100,130,160,200,250,500), labels = c("A+","A","B","C","D","E","F","G","H"))

# Control variable for the upload date of the observation (monthly control).
SalesPrice$dateUpload <- SalesPrice$firstSeenDay
SalesPrice$dateUpload <- as.Date.character(SalesPrice$dateUpload)
SalesPrice$dateUpload <- format(as.Date(SalesPrice$dateUpload), "%Y-%m")
# Deletion of all observations that were uploaded before January 2019.
SalesPrice <- SalesPrice[(SalesPrice$dateUpload>"2018-12"),]
```

```
# N = 32 371

# There are no observations with an upload date later than 12/2020 present in the data set.

# Deletion of any left-over incomplete cases
SalesPrice <- SalesPrice[complete.cases(SalesPrice[ , 1:20]),]
# N = 31 426

## Plots and descriptive statistics

# Plot EPC levels
# Not the plot used in the final thesis document. Just used as an overview.
plot(SalesPrice$energie.klasse.berechnet)
# Data used for the bar plot shown in Figure 3 in final thesis document.
summary(SalesPrice$energie.klasse.berechnet)

# Descriptive Statistics
summary(SalesPrice)
print(summary(SalesPrice$preise.preisValues.Kaufpreis), digits=20)
sd(SalesPrice$preise.preisValues.Kaufpreis)
sd(SalesPrice$energieausweis.energieKennwerte.ENERGIEKENNWERT.kennwert)
sd(SalesPrice$flÃ.chen.flÃ.cheValues.Wohnflaeche)
sd(SalesPrice$anzahlen.ZIMMER)
sd(SalesPrice$saniert)
sd(SalesPrice$erstbezug)
sd(SalesPrice$denkmalschutz)
sd(SalesPrice$aufzug)
sd(SalesPrice$stellplatz)
sd(SalesPrice$jahreszahlen.BAUJAHR)
sd(SalesPrice$vermietet)
sd(SalesPrice$preise.provisionsfrei)

# Calculation of the correlation matrix
SalesPriceCor <- SalesPrice[,c(2:11, 13, 14)]
SalesPriceCor <- SalesPriceCor[,c(11, 4, 6, 1,8 , 5, 3, 2, 9, 7, 10,12)]
SalesPriceCorMatrix <- cor(SalesPriceCor)
# Export as CSV file.
write.csv(SalesPriceCorMatrix, file = "SalesPriceCorMatrix.csv")

## Sales price linear model

# Change class of variables to factor
SalesPrice$aufzug <- as.factor(SalesPrice$aufzug)
SalesPrice$denkmalschutz <- as.factor(SalesPrice$denkmalschutz)
SalesPrice$erstbezug <- as.factor(SalesPrice$erstbezug)
SalesPrice$saniert <- as.factor(SalesPrice$saniert)
SalesPrice$stellplatz <- as.factor(SalesPrice$stellplatz)
SalesPrice$vermietet <- as.factor(SalesPrice$vermietet)
SalesPrice$preise.provisionsfrei <- as.factor(SalesPrice$preise.provisionsfrei)
SalesPrice$addr.plz <- as.factor(SalesPrice$addr.plz)
SalesPrice$dateUpload <- as.factor(SalesPrice$dateUpload)
SalesPrice$anzahlen.ZIMMER <- as.factor(SalesPrice$anzahlen.ZIMMER)

# Change the reference value of the EPC level to D
SalesPrice$energie.klasse.berechnet <- relevel(SalesPrice$energie.klasse.berechnet, ref = "D")

# Definition of the multiple linear regression model
LinearModelSalesPrice <- lm(formula = log(SalesPrice$preise.preisValues.Kaufpreis) ~
                SalesPrice$energie.klasse.berechnet +
                log(SalesPrice$flÃ.chen.flÃ.cheValues.Wohnflaeche) +
                SalesPrice$anzahlen.ZIMMER +
                SalesPrice$saniert +
                SalesPrice$erstbezug +
                SalesPrice$denkmalschutz +
                SalesPrice$aufzug +
                SalesPrice$preise.provisionsfrei +
                SalesPrice$vermietet +
                SalesPrice$stellplatz +
                SalesPrice$objektTyp +
                SalesPrice$baujahr.abschnitte +
                SalesPrice$dateUpload +
                SalesPrice$addr.plz)

# Summary of the Sales Price Model
summary(LinearModelSalesPrice)

# Plots of Sales Price Model to check assumptions
# Package:{graphics}
plot(LinearModelSalesPrice)

# Calculation of robust standard errors
# Using coeftest() from the package {lmtest} and the heteroscedasticity-consistent covariance
matrix estimation vcovHC() from the package {sandwich}. The estimator used is "HC0" that
gives White's estimator. (Package {zoo} needed for package {lmtest}.)
library(lmtest)
library(sandwich)
coeftest(LinearModelSalesPrice, vcov = vcovHC(LinearModelSalesPrice, type = "HC0"))

# Exogeneity-Assumption: The mean of the error term is equal to zero.
mean(LinearModelSalesPrice$residuals)

# Autocorrelation-Assumption: There exists no covariance between error terms.
# Using durbinWatsonTest() from the {car} package for computation.
library(car)
durbinWatsonTest(LinearModelSalesPrice)

# Multicollinearity-Assumption: There exists no perfect multicollinearity between explanatory
variables.
# Using vif() from the {car} package for computation.
vif(LinearModelSalesPrice)

## END OF R - SCRIPT FOR SALES PRICE ANALYSIS ##
```

Appendix 16: Telephone call log with Roland Weber

Telephone Call between Green-Towers Sustainable High-Rises GmbH CEO Roland Weber and Timo Deller, author of this thesis.
Date: 18.01.2021
Topic: Plausibility check of data used for analysis.

Start of log.

- Deller stated that it is important to perform a plausibility check regarding the Data used and suggested going through his statements used in the R code.

- Weber gave the following remarks regarding some of these statements:
    - Energy Efficiency:
        - Buildings with energy consumption of 5 kwh / sqm and annum or less rarely exist. Most likely, wrong information was entered.
        - Buildings with energy consumption of 500 kwh / sqm and annum or more point towards structural deficits of the building and seem unplausible.
    - Building size:
        - It is difficult to define clear cut-off points for such measures.
        - A maximum of 100 m^2 per room seems reasonable.
        - A minimum (because of construction constraints and because bathroom and kitchen are not counted) of 8 m^2 per room seems reasonable.
    - Operating costs:
        - It is difficult to define clear cut-off points for such measures.
        - Not including data with values of 10% or below and 50% or above of cold rent as operational costs should exclude most unplausible information entered.

End of log.