

Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Национальный исследовательский ядерный университет «МИФИ»

## ОТЧЕТ

экзамена по дисциплине: «Эконометрика»

Выполнил: студент группы Б20-902 Калина В.Г.

---

(подпись) (Фамилия И.О.)

Проверил: Смирнов Д. С.

---

(оценка) (подпись)

(Фамилия И.О.)

Москва 2023 г

# Содержание

<b>Содержание</b> .....	<b>2</b>
<b>Цели и задачи</b> .....	<b>3</b>
<b>Описание данных</b> .....	<b>4</b>
Источники.....	4
Статистический анализ.....	4
<b>Модели и тесты</b> .....	<b>9</b>
Модель множественной регрессии.....	9
Мультиколлинеарность.....	10
Алгоритм Фаррара-Глобера.....	10
Метод инфляции дисперсии факторов.....	10
Устранение методом исключения факторов.....	11
Устранение методом добавления факторов.....	12
LASSO-регрессия.....	14
Ridge-регрессия.....	14
Гетероскедастичность.....	15
Тест Бройша-Пагана.....	15
Тест Голдфелда - Куандта.....	15
Тест Глейзера.....	15
Устранение методом исключения факторов.....	15
Устранение методом взвешенного МНК.....	15
Тест Дарбина-Уотсона.....	16
Процедура Кохрейна-Оркатта.....	17
Объединение прогнозов.....	17
<b>Выводы</b> .....	<b>18</b>

## Цели и задачи

Необходимо найти и проанализировать статистические данные.  
Сооздать модели и осуществить прогноз.

1. Найти данные, удовлетворяющие условиям
2. Провести обработку данных
3. Провести статистический анализ
4. Проверить данные на мультиколлинеарность и обучить модели, избавляющиеся от неё
5. Проверить данные на гетероскедастичность и обучить модели, избавляющиеся от неё
6. Объединить наилучшие модели
7. Сделать выводы

# Описание данных

## Источники

Для получения данных был проведён парсинг сайта [cian.ru](https://cian.ru). Были взяты однокомнатные квартиры, сдающиеся в аренду. Таблица имеет следующие атрибуты:

Price - цена за месяц

all\_sqr - общая площадь квартиры

live\_sqr - площадь для жилья (может совпадать с all\_sqr)

kitch - площадь кухни

floor - этаж

pledge - залог

comis - комиссия

looks - количество просмотров публикации

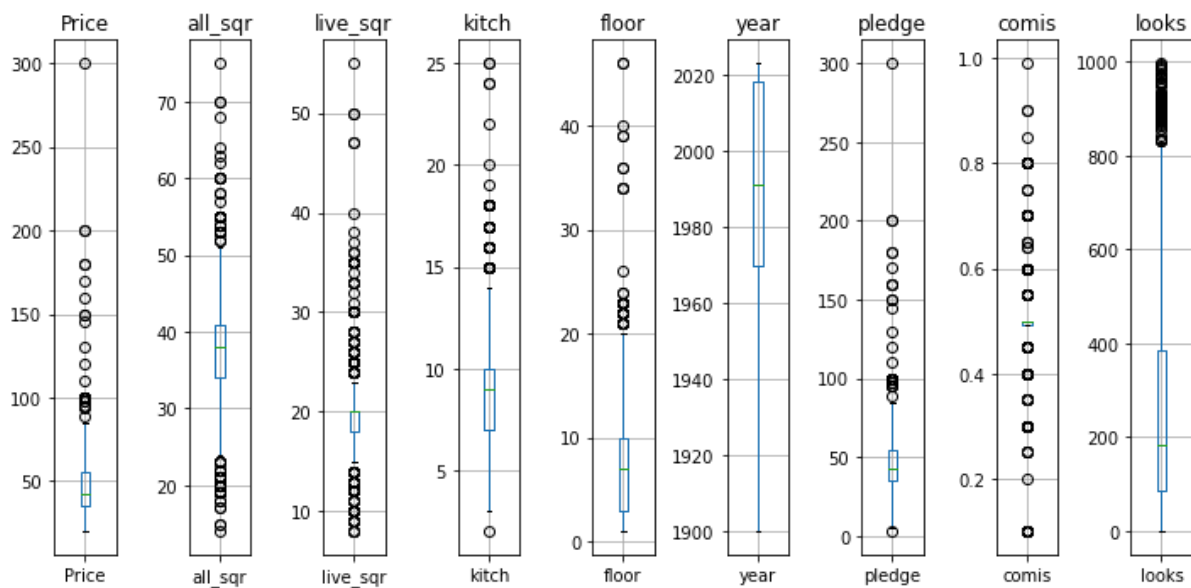
prepayment - предоплата

## Статистический анализ

### Основные характеристики

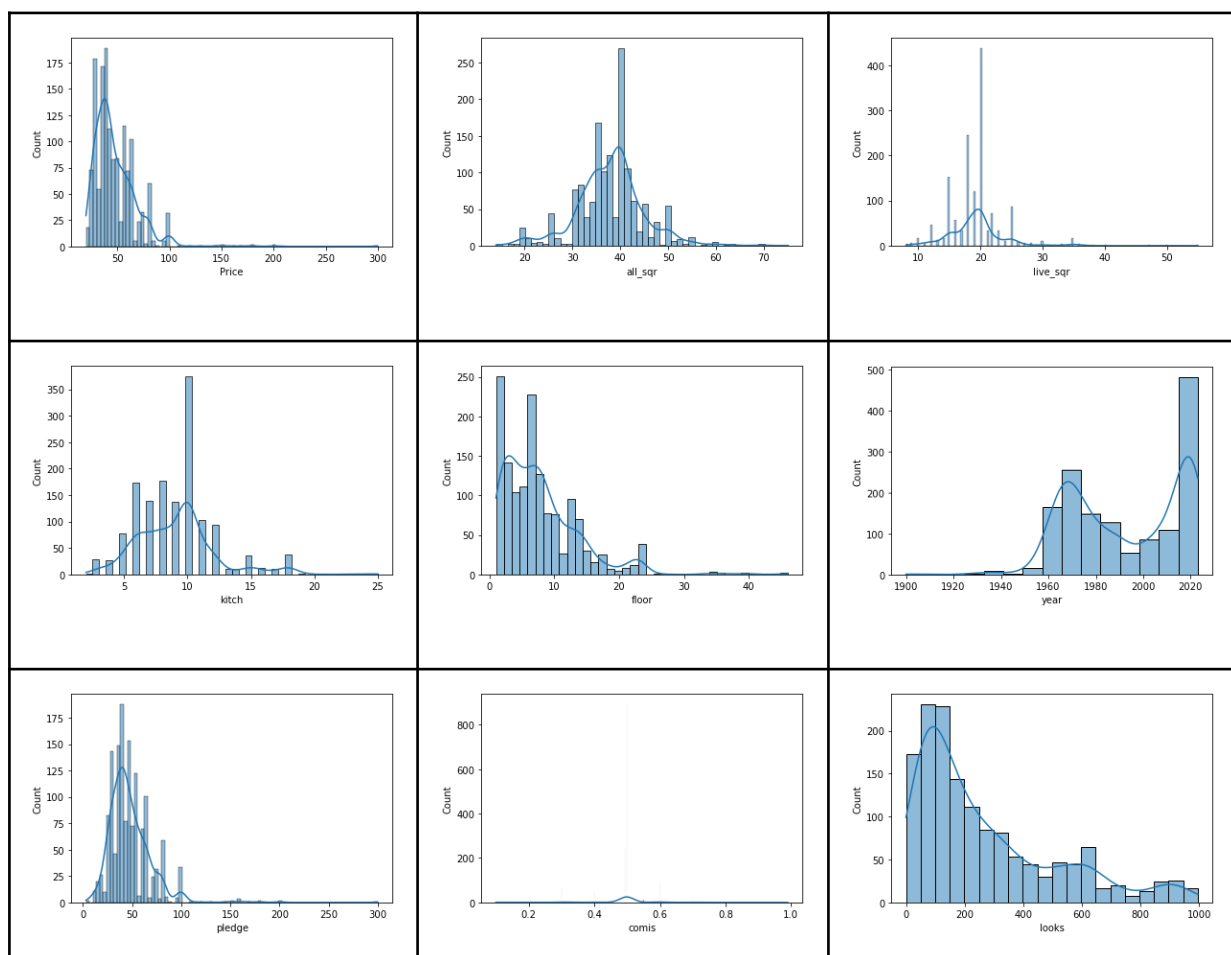
	Mean	Std	ymmetry	Excess	25%	50%	75%	Low	Hight
Price	47.8525	21.0963	3.0067	21.1152	35.0000	42.0000	55.0000	46.9432	48.7619
all_sqr	37.9074	7.0164	0.1755	2.2017	34.0000	38.0000	41.0000	37.6050	38.2098
live_sqr	19.3333	4.4962	1.8024	9.3222	18.0000	20.0000	20.0000	19.1395	19.5271
kitch	9.1296	3.1325	0.9856	2.2964	7.0000	9.0000	10.0000	8.9946	9.2647
floor	7.7558	5.8666	1.7410	5.2295	3.0000	7.0000	10.0000	7.5030	8.0087
year	92.9108	23.6771	-0.1883	-1.2100	70.0000	91.0000	18.0000	91.8902	93.9314
pledge	47.6144	22.0215	2.7567	18.5320	35.0000	43.0000	55.0000	46.6652	48.5637
comis	0.4935	0.0897	-0.8530	7.7703	0.4935	0.5000	0.5000	0.4896	0.4974
looks	67.7558	40.3201	1.2107	0.6075	88.0000	83.0000	84.7500	57.3969	78.1148

### Особенности и выбросы



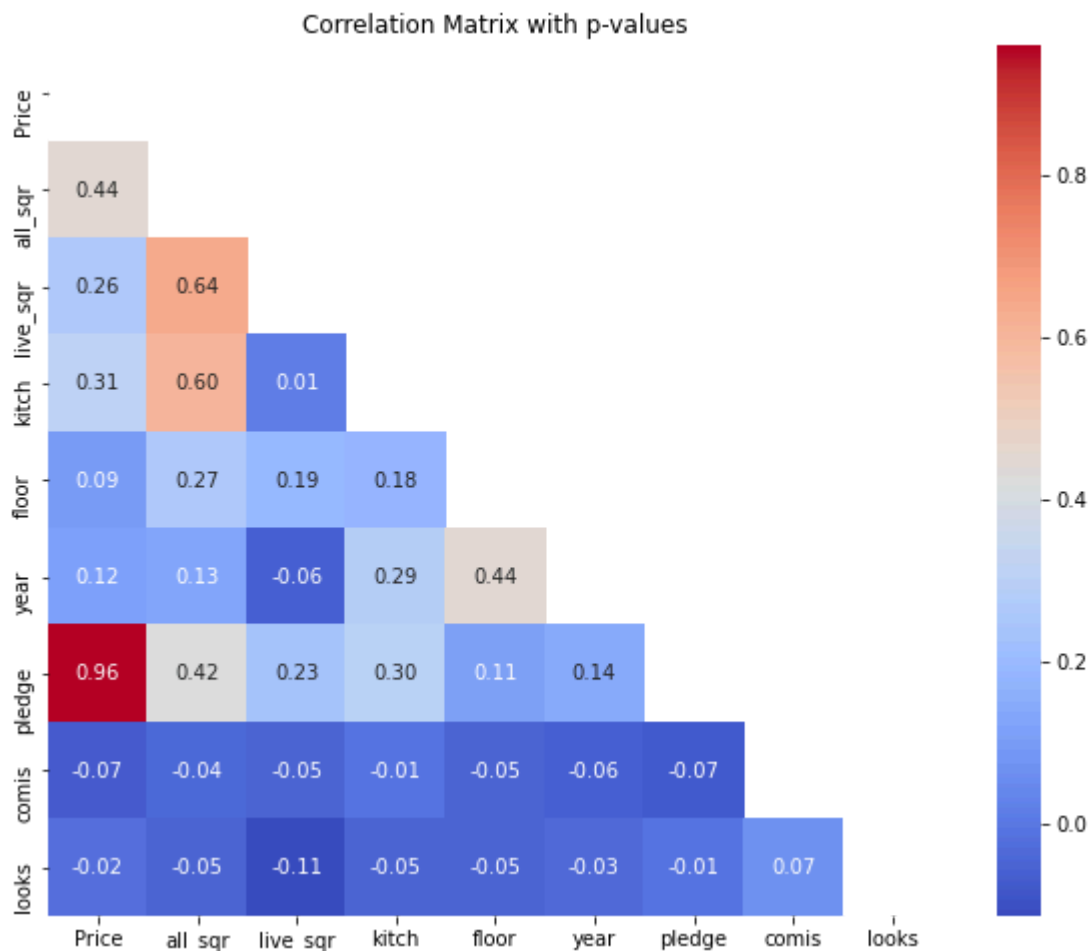
Наименьшие выбросы имеют следующие атрибуты: `kitch`, `year`

### Распределения



Price	all_sqr	live_sqr	kitch	floor	year	pledge	comis	looks
gamma	norm	gamma	gamma	gamma	-	norm	beta	gamma

## Корреляция и её значимость



## Correlation Matrix:

```

Price all_sqr live_sqr kitch floor year pledge comis \
Price 1.0000 0.3942 0.1251 0.3512 0.0968 0.1873 0.9365 -0.2960
all_sqr 0.3942 1.0000 0.5282 0.7285 0.3338 0.1389 0.3585 -0.0942
live_sqr 0.1251 0.5282 1.0000 0.2034 0.1396 -0.1501 0.0809 -0.0163
kitch 0.3512 0.7285 0.2034 1.0000 0.2908 0.3004 0.3346 -0.0544
floor 0.0968 0.3338 0.1396 0.2908 1.0000 0.4592 0.1203 -0.0193
year 0.1873 0.1389 -0.1501 0.3004 0.4592 1.0000 0.2329 -0.1374
pledge 0.9365 0.3585 0.0809 0.3346 0.1203 0.2329 1.0000 -0.3318
comis -0.2960 -0.0942 -0.0163 -0.0544 -0.0193 -0.1374 -0.3318 1.0000
looks -0.0147 -0.0286 -0.1276 -0.0708 -0.0259 -0.0284 0.0014 0.0969

looks
Price -0.0147
all_sqr -0.0286
live_sqr -0.1276
kitch -0.0708
floor -0.0259
year -0.0284
pledge 0.0014
comis 0.0969

```

looks        1.0000

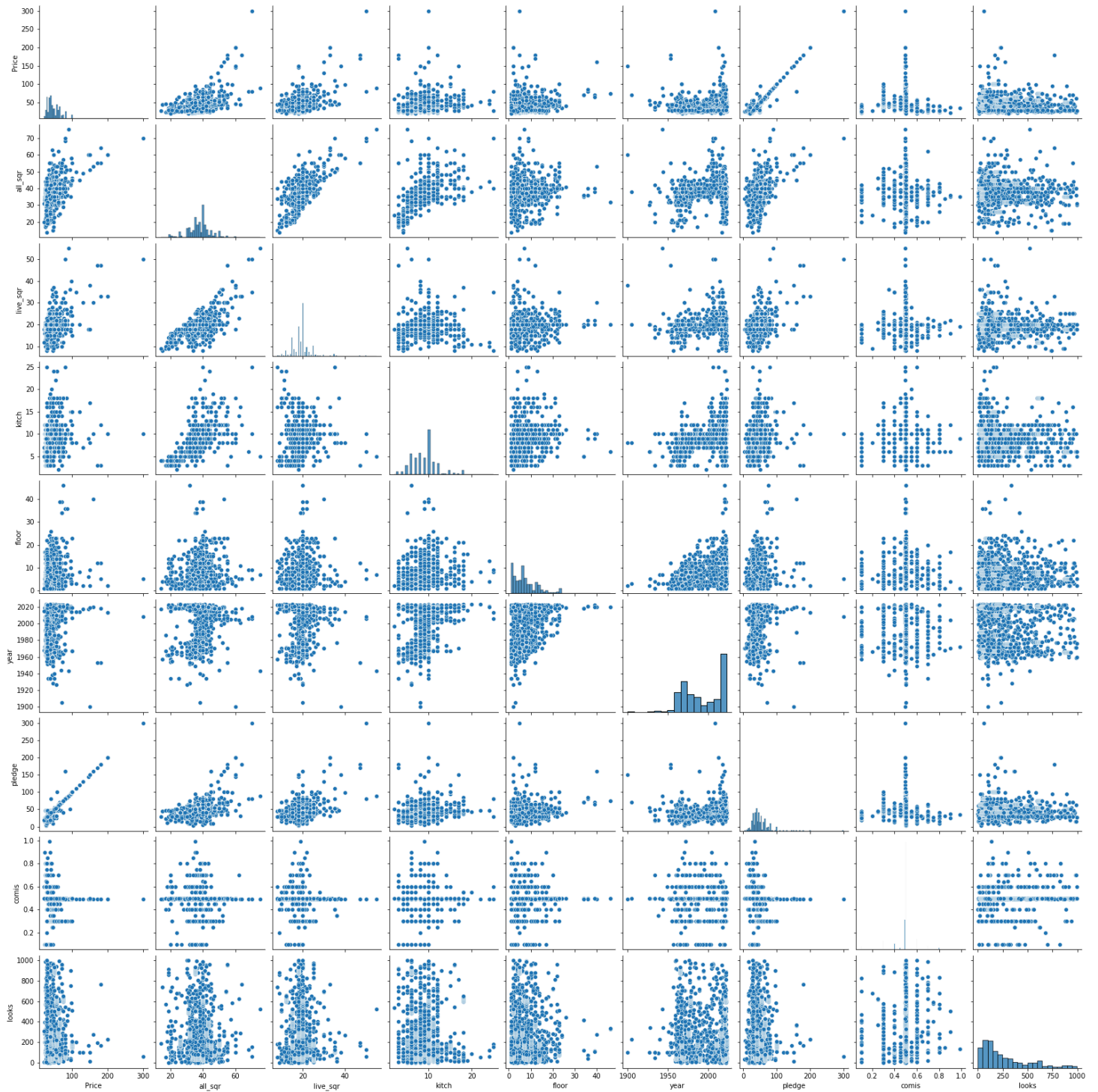
*P-Values Matrix:*

	Price	all_sqr	live_sqr	kitch	floor	year	pledge	comis	looks
Price	0.0000	0.0000	0.0000	0.0000	0.0004	0.0000	0.0000	0.0066	0.4430
all_sqr	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1137	0.0534
live_sqr	0.0000	0.0000	0.0000	0.5975	0.0000	0.0276	0.0000	0.0576	0.0000
kitch	0.0000	0.0000	0.5975	0.0000	0.0000	0.0000	0.0000	0.6881	0.0624
floor	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0657	0.0476
year	0.0000	0.0000	0.0276	0.0000	0.0000	0.0000	0.0000	0.0258	0.2277
pledge	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0049	0.6844
comis	0.0066	0.1137	0.0576	0.6881	0.0657	0.0258	0.0049	0.0000	0.0093
looks	0.4430	0.0534	0.0000	0.0624	0.0476	0.2277	0.6844	0.0093	0.0000

Наибольшая корреляция существует между следующими признаками:  
price - pledge, kitch - all\_sqr, live\_sqr - all\_sqr, однако  
значимыми являются почти все с pledge (кроме kitch, all\_sqr) и  
looks в соотношении с all\_sqr и floor. Большая значимость у таких  
признаков, как live\_sqr - year, Price - floor

## Парная матрица

На следующем графике мы видим подтверждения выводам по корреляциям





# Модели и тесты

## Модель множественной регрессии

Adjusted Coefficients:

const	35.0532
all_sqr	0.1637
live_sqr	0.0884
kitch	0.0600
floor	-0.0752
year	-0.0180
pledge	0.8692
comis	-1.1371
looks	-0.0003

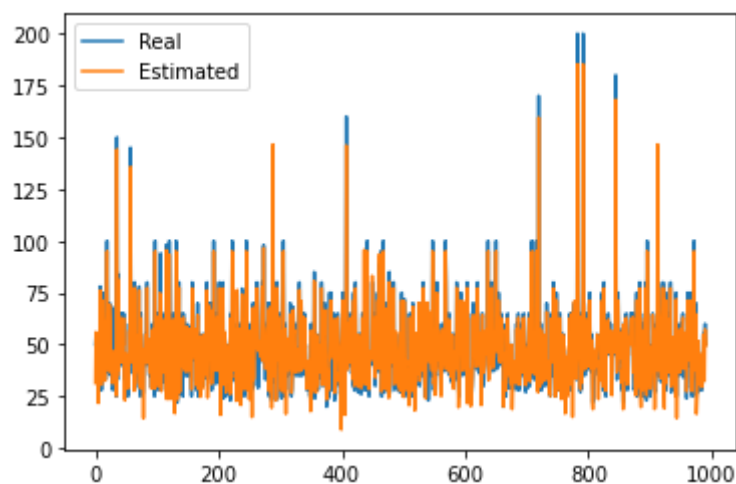
### OLS Regression Results

Dep. Variable:	Price	R-squared:	0.904
Model:	OLS	Adj. R-squared:	0.903
Method:	Least Squares	F-statistic:	1151.
Date:	Tue, 13 Jun 2023	Prob (F-statistic):	0.00
Time:	04:50:38	Log-Likelihood:	-3216.0
No. Observations:	991	AIC:	6450.
Df Residuals:	982	BIC:	6494.
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
--	------	---------	---	------	--------	--------

Остаточная дисперсия 38.922537558208475

Коэффициент детерминации 0.9036627200735821



Классическая модель имеет детерминации множественной регрессии методом МНК уже имеет высокий коэффициент детерминации.

## Мультиколлинеарность

### Алгоритм Фаррара-Глобера

#### Overall Multicollinearity Diagnostics

	MC Results	detection
Determinant $ X'X $ :	0.1254	0
Farrar Chi-Square:	3017.4602	1
Red Indicator:	0.2398	0
Sum of Lambda Inverse:	15.9270	0
Theil's Method:	-3.6659	0
Condition Number:	392.9988	1

1 --> COLLINEARITY is detected by the test  
0 --> COLLINEARITY is not detected by the test

Алгоритм Фаррара-Глобера определил мультиколлинеарность в данных

### Метод инфляции дисперсии факторов

Variable	VIF
Price	80.5311
all_sqr	135.9106
live_sqr	54.2873
kitch	24.1197
floor	3.0068
year	66.2167
pledge	72.4637
comis	31.4061
looks	2.3146

Полученные значения VIF можно интерпретировать следующим образом:

VIF 1: отсутствие мультиколлинеарности (независимая переменная не имеет корреляции с другими переменными).

VIF > 1 и < 5: умеренная мультиколлинеарность (некоторая корреляция между независимыми переменными).

VIF > 5: высокая мультиколлинеарность (сильная корреляция между независимыми переменными).

Таким образом, ни у какого признака не отсутствует мультиколлинеарность, умеренная у floor, looks. Остальные признаки имеют высокую мультиколлинеарность.

## Устранение методом исключения факторов

	Variable	VIF
1	live_sqr	19.9597
2	kitch	10.8688
3	floor	2.9791
4	year	67.5417
5	pledge	6.7957
6	comis	34.3311
7	looks	2.2939

	Variable	VIF
1	live_sqr	14.8823
2	kitch	9.4527
3	floor	2.9607
5	pledge	6.7235
6	comis	15.7689
7	looks	2.2350

	Variable	VIF
1	live_sqr	9.2062
2	kitch	8.0443
3	floor	2.9556
5	pledge	6.7149
7	looks	2.0607

	Variable	VIF
2	kitch	6.3252
3	floor	2.7687
5	pledge	5.5873
7	looks	1.9895

	Variable	VIF
3	floor	2.4214
5	pledge	3.0187
7	looks	1.8961

Adjusted Coefficients:

const	5.8514
floor	-0.0477
pledge	0.8930
looks	-0.0007

### OLS Regression Results

```

=====
Dep. Variable:          Price      R-squared:                0.898
Model:                  OLS        Adj. R-squared:             0.898
Method:                 Least Squares    F-statistic:              2908.
Date:                  Mon, 12 Jun 2023    Prob (F-statistic):       0.00
Time:                  18:18:09          Log-Likelihood:           -3242.4
No. Observations:      991            AIC:                     6493.
Df Residuals:          987            BIC:                     6512.
Df Model:              3
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	5.8514	0.585	10.008	0.000	4.704	6.999
floor	-0.0477	0.035	-1.374	0.170	-0.116	0.020
pledge	0.8930	0.010	92.645	0.000	0.874	0.912
looks	-0.0007	0.001	-0.880	0.379	-0.002	0.001

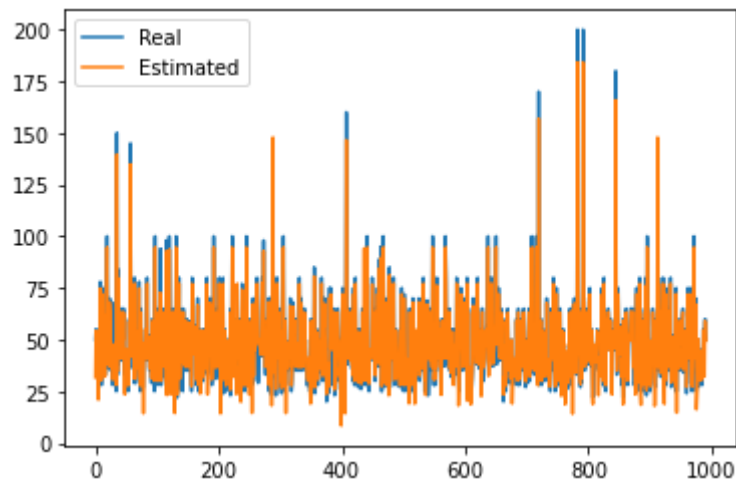
```

=====
Omnibus:                623.229    Durbin-Watson:           1.959
Prob(Omnibus):          0.000      Jarque-Bera (JB):        44034.044
Skew:                   -2.087     Prob(JB):                0.00
Kurtosis:               35.388     Cond. No.                1.04e+03
=====

```

Остаточная дисперсия 40.85041066256296

Коэффициент детерминации 0.8983762246170257



Устранение методом добавления факторов

Adjusted Coefficients:

const	35.0532
pledge	0.8692
all_sqr	0.1637
year	-0.0180
floor	-0.0752

```
live_sqr    0.0884
kitch       0.0600
comis       -1.1371
looks       -0.0003
```

# OLS Regression Results

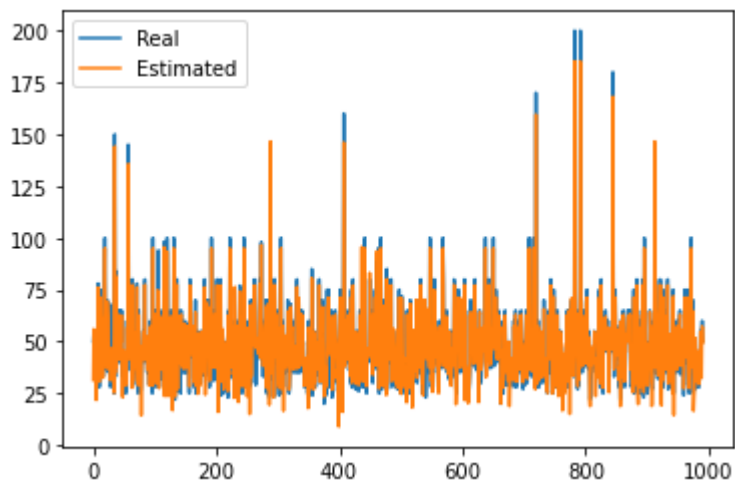
```
=====
Dep. Variable:          Price    R-squared:                0.904
Model:                  OLS      Adj. R-squared:            0.903
Method:                 Least Squares    F-statistic:             1151.
Date:                   Mon, 12 Jun 2023    Prob (F-statistic):       0.00
Time:                   18:20:10    Log-Likelihood:          -3216.0
No. Observations:       991    AIC:                     6450.
Df Residuals:           982    BIC:                     6494.
Df Model:                8
Covariance Type:        nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
-----	-----	-----	-----	-----	-----	-----
const	35.0532	19.245	1.821	0.069	-2.714	72.820
pledge	0.8692	0.010	85.500	0.000	0.849	0.889
all_sqr	0.1637	0.059	2.780	0.006	0.048	0.279
year	-0.0180	0.010	-1.864	0.063	-0.037	0.001
floor	-0.0752	0.039	-1.942	0.052	-0.151	0.001
live_sqr	0.0884	0.072	1.234	0.217	-0.052	0.229
kitch	0.0600	0.101	0.592	0.554	-0.139	0.259
comis	-1.1371	2.326	-0.489	0.625	-5.701	3.427
looks	-0.0003	0.001	-0.362	0.717	-0.002	0.001

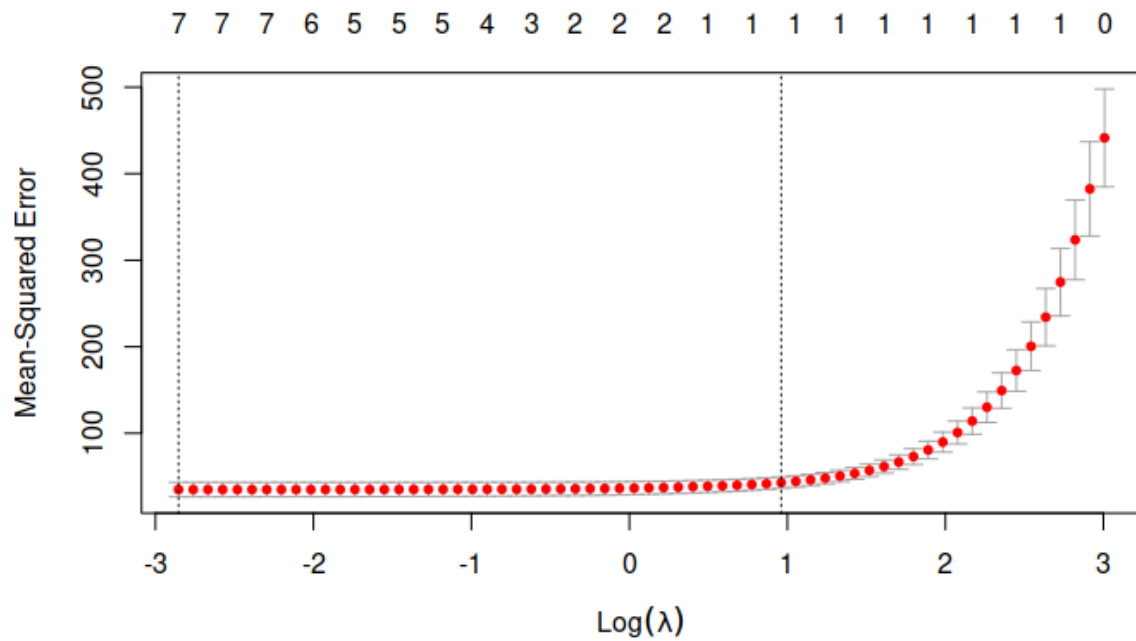
```
=====
Omnibus:                629.498    Durbin-Watson:            1.953
```

Остаточная дисперсия 38.9225375582085

Коэффициент детерминации 0.9036627200735821

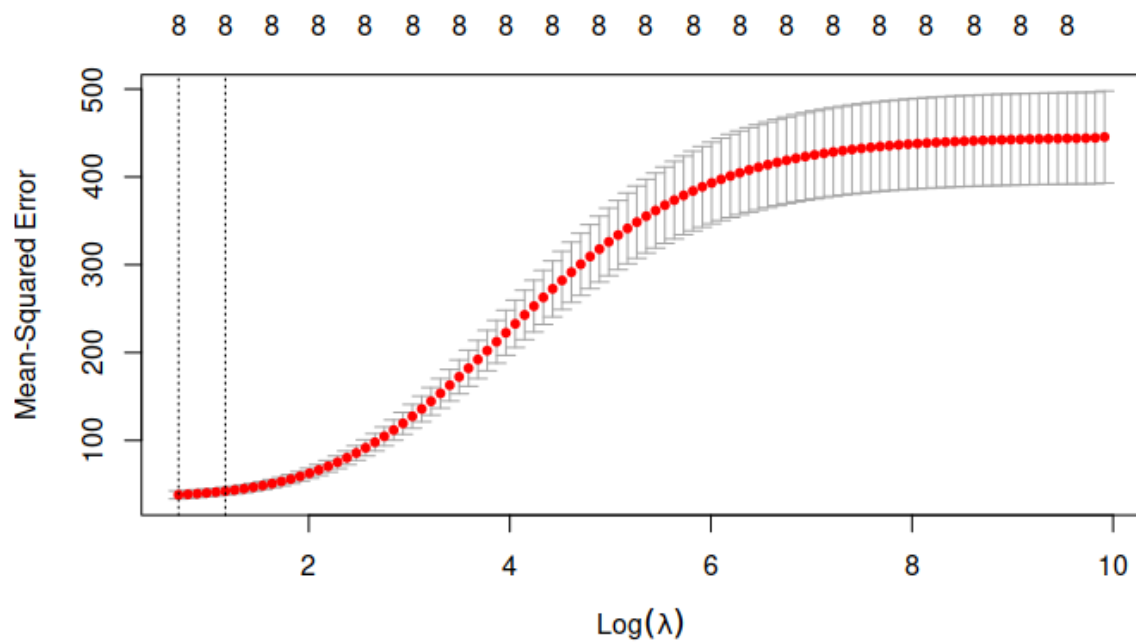


## LASSO-регрессия



Coefficients: [ 1.239 0.265 0.009 -0.334 -0.331 18.416 -0.002  
-0. ]  
Training score: 0.9034919495491971  
Testing score: 0.9585220550275697

## Ridge-регрессия



```
Coefficients: [[ 1.146   0.408   0.187 -0.444 -0.435 18.496 -0.098
-0.073]]
Training score: 0.903662710068247
Testing score: 0.9591180620709993
```

## Гетероскедастичность

Тест Бройша-Пагана

**p-value:** 0.0000003523

Низкое значение p поддерживает наличие гетероскедастичности.

Тест Голдфелда - Куандта

**p-value:** 0.9999999999999999

Если p-значение меньше выбранного уровня значимости, это свидетельствует о гетероскедастичности. Здесь же обратная ситуация

Тест Глейзера

	Estimate	Std. Error	t value	Pr(> t )
(Intercept) == 0	28.5018913	15.0836484	1.890	0.3487
all_sqr == 0	0.1142244	0.0463119	2.466	0.0966 .
live_sqr == 0	0.1012096	0.0568360	1.781	0.4226
kitch == 0	0.0610554	0.0791363	0.772	0.9824
floor == 0	-0.0631180	0.0301995	-2.090	0.2351
year == 0	-0.0148826	0.0075645	-1.967	0.3019
pledge == 0	0.9004356	0.0077205	116.629	<0.001 ***
comis == 0	-0.1242011	1.7169578	-0.072	1.0000
looks == 0	-0.0005853	0.0006466	-0.905	0.9570

Устранение методом исключения факторов

-

Устранение методом взвешенного МНК

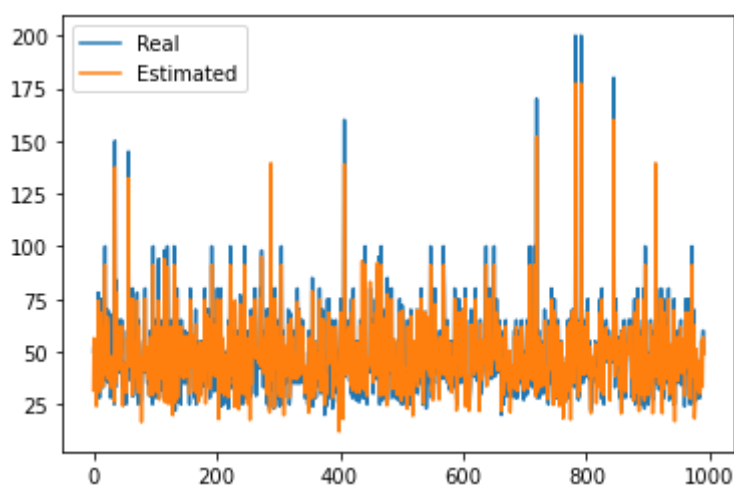
Adjusted Coefficients:

all_sqr	0.2668
live_sqr	0.0244
kitch	-0.1391
floor	-0.1105
year	0.0011
pledge	0.8053
comis	-1.6497
looks	-0.0013

WLS Regression Results						
Dep. Variable:	Price	R-squared (uncentered):	0.982			
Model:	WLS	Adj. R-squared (uncentered):	0.981			
Method:	Least Squares	F-statistic:	6527.			
Date:	Mon, 12 Jun 2023	Prob (F-statistic):	0.00			
Time:	21:53:53	Log-Likelihood:	-3232.8			
No. Observations:	991	AIC:	6482.			
Df Residuals:	983	BIC:	6521.			
Df Model:	8					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
all_sqr	0.2668	0.060	4.451	0.000	0.149	0.384
live_sqr	0.0244	0.078	0.313	0.754	-0.128	0.177
kitch	-0.1391	0.104	-1.337	0.182	-0.343	0.065
floor	-0.1105	0.037	-2.957	0.003	-0.184	-0.037
year	0.0011	0.001	1.358	0.175	-0.000	0.003
pledge	0.8053	0.012	68.621	0.000	0.782	0.828
comis	-1.6497	2.051	-0.805	0.421	-5.674	2.374
looks	-0.0013	0.001	-1.530	0.126	-0.003	0.000
Omnibus:	321.924	Durbin-Watson:	1.966			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3812.176			

Остаточная дисперсия 0.9042477248839821

Коэффициент детерминации 0.9815208385910935



## Тест Дарбина-Уотсона

**Durbin-Watson Statistic:** 1.8853524059917703

*Positive autocorrelation*

- Значение около 2 предполагает отсутствие автокорреляции (остатки независимы).
- Значение меньше 2 предполагает положительную автокорреляцию (остатки положительно коррелируют).



- Значение больше 2 предполагает отрицательную автокорреляцию (остатки имеют отрицательную корреляцию).

Тест выявил, что автокорреляция отсутствует

## Процедура Кохрейна-Оркатта

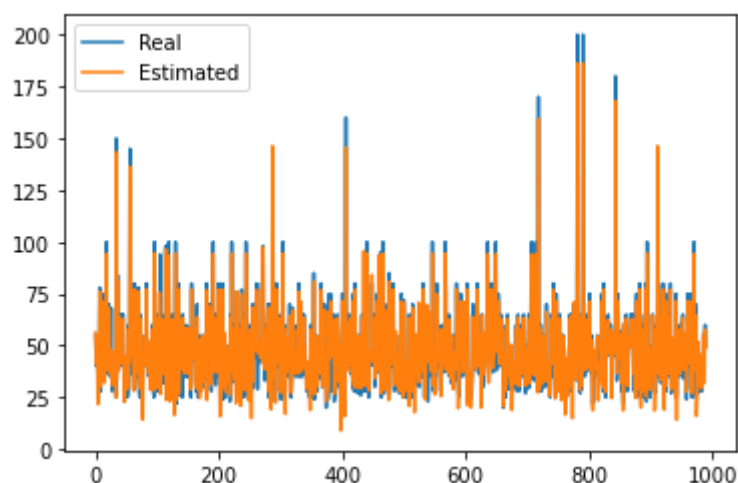
Adjusted Coefficients:

```
all_sqr      0.1753
live_sqr     0.0985
kitch        0.0223
floor        -0.1099
year          -0.0003
pledge       0.8686
comis        -0.8581
looks        -0.0003
lagged       -0.0084
```

GLS Regression Results			
Dep. Variable:	Price	R-squared (uncentered):	0.986
Model:	GLS	Adj. R-squared (uncentered):	0.985
Method:	Least Squares	F-statistic:	7474.
Date:	Mon, 12 Jun 2023	Prob (F-statistic):	0.00
Time:	22:10:46	Log-Likelihood:	-3209.8
No. Observations:	990	AIC:	6438.
Df Residuals:	981	BIC:	6482.
Df Model:	9		
Covariance Type:	nonrobust		

Остаточная дисперсия 38.69336523246564

Коэффициент детерминации 0.9856262529184153



## Объединение прогнозов

-

## Выводы

Были проведены тесты и на мультиколлинеарность, и на гетероскедастичность. Как оказалось, данные имеют высокую мультиколлинеарность. Модель устранения методом добавления факторов отразила наилучший прогноз (90.4%). Тесты на гетероскедастичность показали разные результаты. Модель с процедурой Кохрейна-Оркатта оказалась наиболее точной в прогнозе (98.6%). Схожий результат показала модель устранения гетероскедастичности методом взвешенного МНК.