Regression analysis on house prices

May 7, 2020

1 Introduction

In this project, we will use statistical methods to predict the house prices in Ames, Iowa. The Ames Housing dataset was compiled by Dean De Cock for use in data science education. The goal of this study is to build a regression model to predict the house price using the multiple features from the house price data.

This report is organized as follows: - In Section 2, we will describe the data set and do some exploratory data analysis; - In Section 3, we will build a linear regression model and conduct model selection with forward selection method; - In Section 4, we will add interaction terms to the linear model and conduct model selection with forward selection method; - In Section 5, we will do some diagnositic tests and influence analysis; - Finally, in Section 6, we will summarize this results.

Each model's result is summarized in the Appendix.

2 Data description

The data set analyzed in this project is from Kaggle competition. Let's first take a look at the data.

[1]:
```
# Set up environment
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
/kaggle/input/house-prices-advanced-regression-techniques/train.csv
/kaggle/input/house-prices-advanced-regression-techniques/test.csv
/kaggle/input/house-prices-advanced-regression-techniques/sample_submission.csv
/kaggle/input/house-prices-advanced-regression-techniques/data_description.txt
```

2.1 load libraries and data

[2]:
```
%matplotlib inline
import matplotlib.pyplot as plt
import scipy.stats as stats
import sklearn.linear_model as linear_model
import seaborn as sns
import xgboost as xgb
from sklearn.model_selection import KFold
from IPython.display import HTML, display
from sklearn.manifold import TSNE
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
```

[369]:
```python
# Load data
all_data = pd.read_csv('../input/house-prices-advanced-regression-techniques/
↪train.csv')
all_data.head()
```

[369]:
```
   Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  \
0   1          60       RL         65.0     8450   Pave   NaN      Reg
1   2          20       RL         80.0     9600   Pave   NaN      Reg
2   3          60       RL         68.0    11250   Pave   NaN      IR1
3   4          70       RL         60.0     9550   Pave   NaN      IR1
4   5          60       RL         84.0    14260   Pave   NaN      IR1

  LandContour Utilities  … PoolArea PoolQC Fence MiscFeature MiscVal MoSold  \
0         Lvl    AllPub  …        0    NaN   NaN         NaN       0      2
1         Lvl    AllPub  …        0    NaN   NaN         NaN       0      5
2         Lvl    AllPub  …        0    NaN   NaN         NaN       0      9
3         Lvl    AllPub  …        0    NaN   NaN         NaN       0      2
4         Lvl    AllPub  …        0    NaN   NaN         NaN       0     12

  YrSold  SaleType  SaleCondition  SalePrice
0   2008        WD         Normal     208500
1   2007        WD         Normal     181500
2   2008        WD         Normal     223500
3   2006        WD        Abnorml     140000
4   2008        WD         Normal     250000

[5 rows x 81 columns]
```

There are 79 explanatory variables describing (almost) every aspect of residential homes, including both categorical and numerical variables: - MSSubClass: the type of dwelling involved in the sale - MSZoning: Identifies the general zoning classification of the sale. - LotFrontage: Linear feet of street connected to property - LotArea: Lot size in square feet - Street: Type of road access to property - Alley: Type of alley access to property - LotShape: General shape of property - LandContour: Flatness of the property - Utilities: Type of utilities available - LotConfig: Lot configuration - LandSlope: Slope of property - Neighborhood: Physical locations within Ames city limits - Condition1: Proximity to various conditions - Condition2: Proximity to various conditions (if more than one is present) - BldgType: Type of dwelling - HouseStyle: Style of dwelling - OverallQual: Rates the overall material and finish of the house - OverallCond: Rates the overall condition of the house - YearBuilt: Original construction date - YearRemodAdd: Remodel date (same as construction date if no remodeling or additions) - RoofStyle: Type of roof - RoofMatl: Roof material - Exterior1st: Exterior covering on house - Exterior2nd: Exterior covering on house (if more than one material) - MasVnrType: Masonry veneer type - MasVnrArea: Masonry veneer area in square feet - ExterQual: Evaluates the quality of the material on the exterior - ExterCond: Evaluates the present condition of the material on the exterior - Foundation: Type of foundation - BsmtQual: Evaluates the height of the basement - BsmtCond: Evaluates the general condition of the basement - BsmtExposure: Refers to walkout or garden level walls - BsmtFinType1: Rating of basement finished area - BsmtFinSF1: Type 1 finished square feet - BsmtFinType2: Rating of basement finished area (if multiple types) - BsmtFinSF2: Type 2 finished square feet - BsmtUnfSF: Unfinished square feet of basement area - TotalBsmtSF: Total square feet of basement area - Heating: Type of heating - HeatingQC: Heating quality and condition - CentralAir: Central air conditioning - Electrical: Electrical system - 1stFlrSF: First Floor square feet - 2ndFlrSF: Second floor square feet - LowQualFinSF: Low quality finished square feet (all floors) - GrLivArea: Above grade (ground) living area square feet - BsmtFullBath: Basement full bathrooms - BsmtHalfBath: Basement half bathrooms - FullBath: Full bathrooms above grade - HalfBath: Half baths above grade - Bedroom: Bedrooms above grade (does NOT include basement bedrooms) - Kitchen: Kitchens above grade - KitchenQual: Kitchen quality - TotRmsAbvGrd: Total rooms above grade (does not include bathrooms) - Functional: Home functionality (Assume typical unless deductions are warranted) - Fireplaces: Number of fireplaces - FireplaceQu: Fireplace quality - GarageType: Garage location - GarageYrBlt: Year garage was built - GarageFinish: Interior finish of the garage - GarageCars: Size of garage in car capacity - GarageArea: Size of garage in square feet - GarageQual: Garage quality - GarageCond: Garage condition - PavedDrive: Paved driveway - WoodDeckSF: Wood deck area in square feet - OpenPorchSF: Open porch area in square feet - EnclosedPorch: Enclosed porch area in square feet - 3SsnPorch: Three season porch area in square feet - ScreenPorch: Screen porch area in square feet - PoolArea: Pool area in square feet - PoolQC: Pool quality - Fence: Fence quality - MiscFeature: Miscellaneous feature not covered in other categories - MiscVal: Value of miscellaneous feature - MoSold: Month Sold (MM) - YrSold: Year Sold (YYYY) - SaleType: Type of sale - SaleCondition: Condition of sale.

The dependent variable is the house price ("SalePrice").

It can be seen from the data frame that some data are missing. Before we get started with modeling, we need to clean the data frame and conduct explorotary data analysis.

2.2 Missing data

[370]:
```
# Missing data
total = all_data.isnull().sum().sort_values(ascending=False)
percent = (all_data.isnull().sum()/all_data.isnull().count()).
 ↪sort_values(ascending=False)
missing_data = pd.concat([total, percent], axis=1, keys=['Total', 'Percent'])
missing_data.head(20)
```

[370]:

	Total	Percent
PoolQC	1453	0.995205
MiscFeature	1406	0.963014
Alley	1369	0.937671
Fence	1179	0.807534
FireplaceQu	690	0.472603
LotFrontage	259	0.177397
GarageCond	81	0.055479
GarageType	81	0.055479
GarageYrBlt	81	0.055479
GarageFinish	81	0.055479
GarageQual	81	0.055479
BsmtExposure	38	0.026027

```
BsmtFinType2     38  0.026027
BsmtFinType1     37  0.025342
BsmtCond         37  0.025342
BsmtQual         37  0.025342
MasVnrArea        8  0.005479
MasVnrType        8  0.005479
Electrical        1  0.000685
Utilities         0  0.000000
```

The above table shows the percentage of missing data in each column. For the columns with over 15% missing data, we drop the whole columns, i.e. the features. For the rest missing data, we could delete the rows, i.e., the samples.

[371]:
```
data = all_data.
↪drop(['PoolQC','MiscFeature','Alley','Fence','FireplaceQu','LotFrontage','Id'],axis␣
↪= 1)
data = data.dropna()
```

[372]:
```
data.describe()
```

[372]:
```
         MSSubClass        LotArea  OverallQual  OverallCond    YearBuilt  \
count  1338.000000    1338.000000  1338.000000  1338.000000  1338.000000
mean     56.136024   10706.294469     6.219731     5.596413  1973.029148
std      41.252576   10336.621126     1.324472     1.078124    29.563540
min      20.000000    1300.000000     2.000000     2.000000  1880.000000
25%      20.000000    7744.000000     5.000000     5.000000  1956.000000
50%      50.000000    9600.000000     6.000000     5.000000  1976.000000
75%      70.000000   11760.750000     7.000000     6.000000  2001.000000
max     190.000000  215245.000000    10.000000     9.000000  2010.000000

       YearRemodAdd   MasVnrArea   BsmtFinSF1   BsmtFinSF2    BsmtUnfSF  …  \
count   1338.000000  1338.000000  1338.000000  1338.000000  1338.000000  …
mean    1985.668909   110.360239   464.234679    49.218236   582.494768  …
std       20.296463   185.604816   458.792420   166.196584   439.950528  …
min     1950.000000     0.000000     0.000000     0.000000     0.000000  …
25%     1968.000000     0.000000     0.000000     0.000000   248.000000  …
50%     1994.500000     0.000000   413.000000     0.000000   489.000000  …
75%     2004.000000   174.000000   733.000000     0.000000   815.750000  …
max     2010.000000  1600.000000  5644.000000  1474.000000  2336.000000  …

         WoodDeckSF  OpenPorchSF  EnclosedPorch    3SsnPorch  ScreenPorch  \
count  1338.000000  1338.000000    1338.000000  1338.000000   1338.00000
mean     99.384903    47.782511      21.263827     3.585949     16.43423
std     127.537065    65.362562      60.843964    30.224622     58.05159
min       0.000000     0.000000       0.000000     0.000000      0.00000
25%       0.000000     0.000000       0.000000     0.000000      0.00000
50%       6.000000    28.000000       0.000000     0.000000      0.00000
```

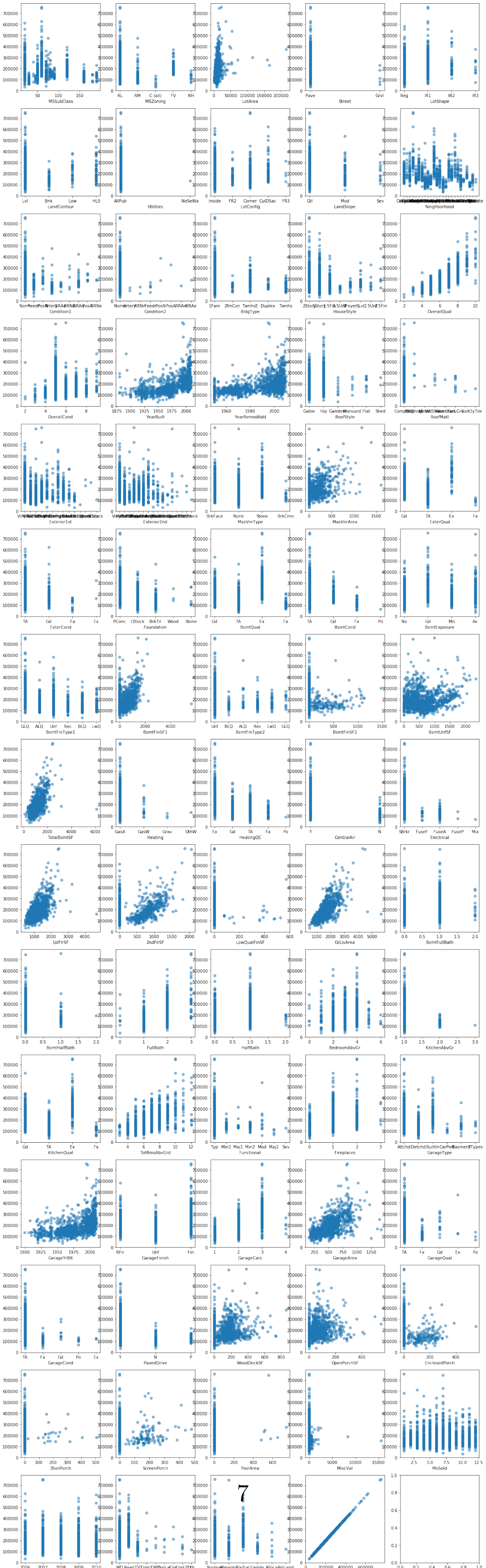
```
75%      174.500000    70.000000      0.000000    0.000000      0.00000
max      857.000000   547.000000    552.000000  508.000000    480.00000

            PoolArea       MiscVal        MoSold       YrSold      SalePrice
count    1338.000000   1338.000000   1338.000000  1338.000000    1338.000000
mean        3.010463     42.932735      6.331839  2007.805680  186761.782511
std        41.961337    508.056255      2.699437     1.330691   78913.847668
min         0.000000      0.000000      1.000000  2006.000000   35311.000000
25%         0.000000      0.000000      5.000000  2007.000000  135000.000000
50%         0.000000      0.000000      6.000000  2008.000000  168500.000000
75%         0.000000      0.000000      8.000000  2009.000000  220000.000000
max       738.000000  15500.000000     12.000000  2010.000000  755000.000000

[8 rows x 36 columns]
```

2.3 Preliminary variables selection

In this part, we will use the cleaned data from last part to do some preliminary variables selection. We will drop the outliers and unnecessary variables. To do that, let's plot the scatter matrix of all variables. In the plot, the y-axis is "SalePrice".

2.3.1 Scatter matrix

[363]:
```
fig,axes = plt.subplots(15,5,figsize=(20,70))
axs = axes.ravel()
for i, col in enumerate(data.columns.values):
    ax = axs[i]
    ax.scatter(data[col], data['SalePrice'], alpha=0.5)
    ax.set_xlabel(col)
```



MSSubClass
MSZoning
LotArea
Street
LotShape
LandContour
Utilities
LotConfig
LandSlope
Neighborhood
Condition1
Condition2
BldgType
HouseStyle
OverallQual
OverallCond
YearBuilt
YearRemodAdd
RoofStyle
RoofMatl
Exterior1st
Exterior2nd
MasVnrType
MasVnrArea
ExterQual
ExterCond
Foundation
BsmtQual
BsmtCond
BsmtExposure
BsmtFinType1
BsmtFinSF1
BsmtFinType2
BsmtFinSF2
BsmtUnfSF
TotalBsmtSF
Heating
HeatingQC
CentralAir
Electrical
1stFlrSF
2ndFlrSF
LowQualFinSF
GrLivArea
BsmtFullBath
BsmtHalfBath
FullBath
HalfBath
BedroomAbvGr
KitchenAbvGr
KitchenQual
TotRmsAbvGrd
Functional
Fireplaces
GarageType
GarageYrBlt
GarageFinish
GarageCars
GarageArea
GarageQual
GarageCond
PavedDrive
WoodDeckSF
OpenPorchSF
EnclosedPorch
3SsnPorch
ScreenPorch
PoolArea
MiscVal
MoSold
YrSold
SaleType
SaleCondition
SalePrice

Checking the scatter plot, we can see that:

- The following categorical variables could be dropped because the distribution is highly skewed (in other words, there's no need to have multiple categories): 'Street','Condition2', 'RoofStyle', 'RoofMatl', 'LowQualFinSF', 'PoolArea', 'YrSold', 'MoSold', '3SsnPorch', 'GarageCars', 'GarageQual', 'GarageCond', 'Utilities'. We will drop them all.
- There are ony a few samples having LotArea greater than 50,000, and these samples are far from other samples. Thus we could drop these samples. Similarly, there are samples with 'SalePrice' higher than 500000, 'MiscVal' greater than 5000.

[373]:
```
drop_col = ['Street','Condition2', 'RoofStyle',␣
↪'RoofMatl','LowQualFinSF','PoolArea', 'YrSold','MoSold',
            '3SsnPorch','GarageCars','GarageQual','GarageCond','Utilities']
data = data.drop(columns=drop_col)
data = data.loc[data['LotArea'] < 50000]
data = data.loc[data['SalePrice'] < 500000]
data = data.loc[data['MiscVal'] < 5000 ]
```

2.3.2 Identify highly correlated predictors using VIF

Next, we label the quantitative (numerical) and qualitative (categorical) variables, and turn categorical variables to dummy variables. The dummy variables expand the columns of the data frame.

[374]:
```
quantitative = [f for f in data.columns if data.dtypes[f] != 'object']
quantitative.remove('SalePrice')
qualitative = [f for f in data.columns if data.dtypes[f] == 'object']
data = pd.get_dummies(data, columns= qualitative)
```

Before we start the modeling, one more thing to do is normalize the data.

[375]:
```
from scipy import stats
for col in quantitative:
    norm = stats.zscore(data[col])
    data[col] = norm
data['SalePrice'] = stats.zscore(data['SalePrice'])
```

We will only check the correlation among numerical variables because even if the categorical variable is not associated with other variables in the regression model, the categorical variables will necessarily have high VIFs, if the proportion of cases in the reference category is small.

[376]:
```
from statsmodels.stats.outliers_influence import variance_inflation_factor
import statsmodels.api as sm
exog = sm.add_constant(data[quantitative])
vif = pd.DataFrame()
```

```
vif["VIF Factor"] = [variance_inflation_factor(exog.values, i) for i in␣
↪range(exog.shape[1])]
vif["features"] = exog.columns
corelated = vif.loc[vif['VIF Factor']>5]['features'].values
```

```
/opt/conda/lib/python3.7/site-
packages/statsmodels/stats/outliers_influence.py:193: RuntimeWarning: divide by
zero encountered in double_scalars
  vif = 1. / (1. - r_squared_i)
```

[377]:
```
corelated
```

[377]:
```
array(['YearBuilt', 'BsmtFinSF1', 'BsmtFinSF2', 'BsmtUnfSF',
       'TotalBsmtSF', '1stFlrSF', '2ndFlrSF', 'GrLivArea'], dtype=object)
```

The above variables are correlated, so we will drop them.

[378]:
```
data = data.drop(corelated,axis = 1)
```

3 Linear regression model

In this section,we will develop a linear regression model for the data. Specifically, we will address the following questions:

- Which predictor variables should be included in the model?
- Are the assumptions (constant variance, noramlity) satisfied?

We will first buid a full linear model using all variables. Then we make hypothesis that some variables are unnecessary based on the p-values, and test the hypothesis.

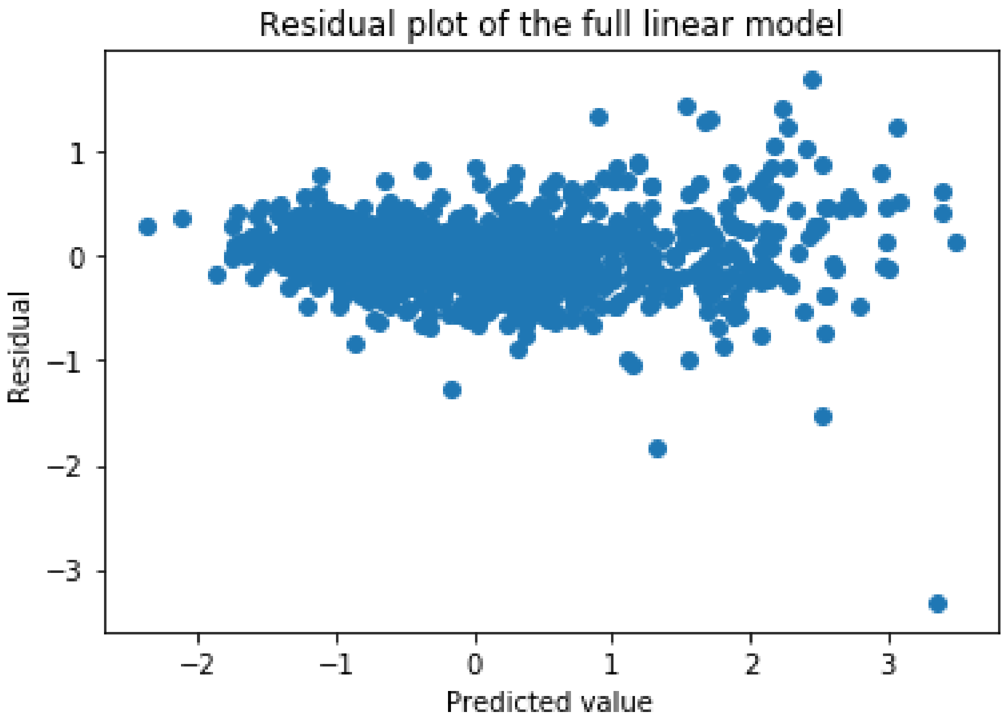
3.1 Full linear model with all variables

[379]:
```
import statsmodels.api as sm
exog = data.drop(['SalePrice'],axis = 1)
basemodel = sm.OLS(data['SalePrice'], exog)
base_results = basemodel.fit()
```

Check the residual plot to find any outliers.

[380]:
```
res = base_results.resid
plt.scatter(base_results.fittedvalues,res)
plt.xlabel("Predicted value")
plt.ylabel("Residual")
plt.title("Residual plot of the full linear model")
```

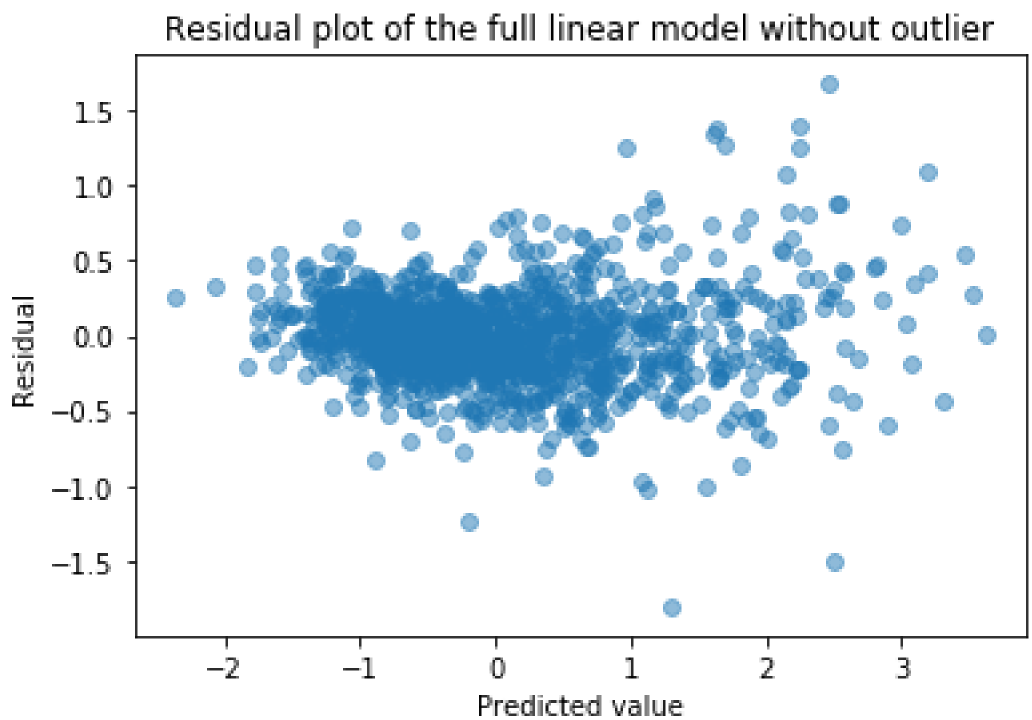
[380]:
```
Text(0.5, 1.0, 'Residual plot of the full linear model')
```



Drop the outlier and we replot the residuals.

[381]:
```
drop_idx = abs(base_results.resid).argmax()
data1 = data.drop(drop_idx,axis = 0).reset_index()
exog = data1.drop(['SalePrice'],axis = 1)
base_results1 = sm.OLS(data1['SalePrice'], exog).fit()
data = data1.drop(drop_idx-1, axis = 0).reset_index()
exog = data.drop(['SalePrice'],axis = 1)
base_results = sm.OLS(data['SalePrice'], exog).fit()
plt.scatter(base_results.fittedvalues,base_results.resid,alpha = 0.5)
plt.xlabel("Predicted value")
plt.ylabel("Residual")
plt.title("Residual plot of the full linear model without outlier")
```

[381]:
```
Text(0.5, 1.0, 'Residual plot of the full linear model without outlier')
```



With dummy variables, there are 181 predictors in total, which is too many. We will drop some unnecessary predictors in the following part.

3.2 Reduced model

In this part, we propose a reduced model by dropping the variables with p-value greater than 0.2 in the full model. Our hypothesis is that all the dropped variables are unnecessary. We will discuss more about the hypothesis in the next part.

[306]:
```
# Drop the variables with p-value greater than 0.05 from the full model
pvalue = pd.DataFrame(base_results.pvalues,columns =['p']).reset_index()
insignificant = pvalue.loc[pvalue['p']>0.2]['index'].values
data_sig = data.drop(insignificant,axis = 1)
data_sig.head()
```

[306]:
```
    LotArea  OverallQual  OverallCond  YearRemodAdd  MasVnrArea  BsmtFullBath  \
0 -0.328162     0.618742    -0.554698      0.858577    0.503894      1.117614
1 -0.074589    -0.151469     2.219496     -0.470149   -0.602333     -0.847537
2  0.289232     0.618742    -0.554698      0.809365    0.311998      1.117614
3 -0.085614     0.618742    -0.554698     -0.765422   -0.602333      1.117614
4  0.952929     1.388954    -0.554698      0.710941    1.373073      1.117614

   FullBath  HalfBath  KitchenAbvGr  TotRmsAbvGrd  …  GarageFinish_RFn  \
0  0.800986  1.194502     -0.170249      0.963374  …                 1
```

```
1  0.800986 -0.796838     -0.170249     -0.334594  …                1
2  0.800986  1.194502     -0.170249     -0.334594  …                1
3 -1.052966 -0.796838     -0.170249      0.314390  …                0
4  0.800986  1.194502     -0.170249      1.612357  …                1

   PavedDrive_P  PavedDrive_Y  SaleType_COD  SaleType_ConLw  SaleType_New  \
0             0             1             0               0             0
1             0             1             0               0             0
2             0             1             0               0             0
3             0             1             0               0             0
4             0             1             0               0             0

   SaleType_WD  SaleCondition_Abnorml  SaleCondition_Alloca  \
0            1                      0                     0
1            1                      0                     0
2            1                      0                     0
3            1                      1                     0
4            1                      0                     0

   SaleCondition_Partial
0                      0
1                      0
2                      0
3                      0
4                      0

[5 rows x 93 columns]
```

[ ]:
```python
# Calculate the p-values of the variables in the first reduced model
exog = sm.add_constant(data_sig.drop(['SalePrice'],axis = 1))
reduced_model = sm.OLS(data_sig['SalePrice'],exog)
reduced_results = reduced_model.fit()
```

Repeat the above process and delete the variables with p-value greater than 0.2.

[319]:
```python
pvalue = pd.DataFrame(reduced_results.pvalues,columns =['p']).reset_index()
insignificant = pvalue.loc[pvalue['p']>0.2]['index'].values
data_sig = data_sig.drop(insignificant[1:],axis = 1)
exog = sm.add_constant(data_sig.drop(['SalePrice'],axis = 1))
reduced_model = sm.OLS(data_sig['SalePrice'],exog)
reduced_results = reduced_model.fit()
```

We will use this model as our final reduced linear model. There are only 73 predictors in the reduced model, while there are 183 in the full model. We dropped 183- 73 = 110 variables in this step!

3.3 Hypothesis test

- Null hypothesis H0: the coefficients of the dropped variables in the last step are all zeros.
- Alternative hypothesis Ha: not all of the coefficients of the 145 variables in the last step are zero.
- Decision rule: we will check the F score, which is defined by F = (SSE(R) - SSE(F))/(df(R) - df(F))/(SSE(F)/df(F)) ~ F(0.95, (df(R) - df(F)), df(F)). If F > F(0.95, (df(R) - df(F)), df(F)), we reject H0; otherwise we fail to reject.

Now let's test the null hypothesis.

[312]:
```
sse_f = base_results.ssr
df_f = base_results.df_resid
print("SSE of the full model is:", sse_f)
print("DF of the full model error is:",df_f)
```

```
SSE of the full model is: 116.34877755433077
DF of the full model error is: 1131.0
```

[320]:
```
sse_r = reduced_results.ssr
df_r = reduced_results.df_resid
print("SSE of the reduced model is:", sse_r)
print("DF of the reduced model error is:",df_r)
```

```
SSE of the reduced model is: 128.27203661502872
DF of the reduced model error is: 1241.0
```

[321]:
```
# f statistics to test H0
F = (sse_r - sse_f)/(df_r - df_f)/(sse_f/df_f)
print(F)
```

```
1.05366626425934
```

- Conclusion: Since F(0.95, 110, 1131) = 1.25 > F, we conclude that we fail to reject H0.

Therefore, the coefficients of the 110 variables in the last step are all zeros. We could use the reduced linear model with only 73 variables to predict the sale price.

3.4 Model selection using forward selection method

Although we have reduced 110 variables in last section, we could further optimize the linear model using forward selection method.

[323]:
```
import statsmodels.formula.api as smf
def forward_selected(data, response):
#     https://planspace.org/20150423-forward_selection_with_statsmodels/
    """Linear model designed by forward selection.

    Parameters:
    -----------
```

```
    data : pandas DataFrame with all possible predictors and response

    response: string, name of response column in data

    Returns:
    --------
    model: an "optimal" fitted statsmodels linear model
           with an intercept
           selected by forward selection
           evaluated by adjusted R-squared
    """
    remaining = set(data.columns)
    remaining.remove(response)
    selected = []
    current_score, best_new_score = 0.0, 0.0
    while remaining and current_score == best_new_score:

        scores_with_candidates = []
        for candidate in remaining:

            formula = "{} ~ {} + 1".format(response,' + '.join(selected +␣
↪[candidate]))
            model = smf.ols(formula, data)

            res = model.fit()
            score = res.rsquared_adj
            scores_with_candidates.append((score, candidate))
        scores_with_candidates.sort()
        best_new_score, best_candidate = scores_with_candidates.pop()
        if current_score < best_new_score:
            remaining.remove(best_candidate)
            selected.append(best_candidate)
            current_score = best_new_score
    formula = "{} ~ {} + 1".format(response,
                                   ' + '.join(selected))
    model = smf.ols(formula, data).fit()
    return model

# Find the optimal linear model
data_sig.rename(columns = {'3SsnPorch':'SsnPorch', '1stFlrSF':
↪'stFlrSF','2ndFlrSF':'ndFlrSF',
                            'Exterior2nd_Wd Sdng':'Exterior2nd_WdSdng',
                      'Exterior2nd_Wd Shng':'Exterior2nd_WdShng','HouseStyle_1.
↪5Unf':'HouseStyle_1_5Unf',
                       'HouseStyle_2.5Fin':'HouseStyle_2_5Fin','HouseStyle_2.
↪5Unf':'HouseStyle_2_5Unf',
```

```
                        'HouseStyle_1.5Fin':
↪'HouseStyle_1_5Fin','RoofMatl_Tar&Grv':'RoofMatl_TarGrv',
                        'Exterior2nd_Brk Cmn':'Exterior2nd_BrkCmn','MSZoning_C␣
↪(all)':'MSZoning_Call',
                        'Exterior1st_Wd Sdng':'Exterior1st_WdSdng'}, inplace =␣
↪True)
optimal_linear_model = forward_selected(data_sig, 'SalePrice')
print(optimal_linear_model.model.formula)
print(optimal_linear_model.rsquared_adj)
```

```
SalePrice ~ OverallQual + LotArea + GarageArea + BsmtQual_Ex + TotRmsAbvGrd +
BsmtFinType1_Unf + YearRemodAdd + Fireplaces + Neighborhood_NoRidge +
KitchenQual_Ex + BsmtExposure_Gd + Neighborhood_Crawfor + KitchenAbvGr +
Neighborhood_StoneBr + FullBath + Exterior1st_BrkFace + Neighborhood_NridgHt +
SaleType_WD + BsmtFullBath + OpenPorchSF + MasVnrArea + Functional_Typ +
WoodDeckSF + SaleCondition_Abnorml + KitchenQual_Gd + BldgType_Twnhs +
Foundation_BrkTil + Neighborhood_BrkSide + MSZoning_FV + OverallCond +
Condition1_Norm + MSZoning_Call + GarageType_2Types + ScreenPorch +
LotConfig_CulDSac + GarageType_BuiltIn + HouseStyle_SLvl + BsmtFinType1_GLQ +
Exterior2nd_WdShng + ExterQual_Ex + LotConfig_FR2 + BsmtExposure_No +
HouseStyle_SFoyer + SaleCondition_Alloca + MasVnrType_BrkCmn + LotConfig_FR3 +
Neighborhood_NWAmes + HalfBath + Condition1_RRAe + LandContour_Low +
Exterior1st_WdSdng + Functional_Min2 + Neighborhood_Gilbert + BsmtQual_Gd +
Neighborhood_MeadowV + Exterior2nd_CmentBd + Neighborhood_Mitchel +
LandSlope_Sev + Exterior2nd_WdSdng + Neighborhood_NAmes + Neighborhood_Edwards +
MSZoning_RL + Condition1_PosN + MasVnrType_Stone + Exterior1st_Stone +
Neighborhood_Timber + Exterior1st_CemntBd + GarageFinish_RFn + MSZoning_RH +
EnclosedPorch + 1
0.8970698226785163
```

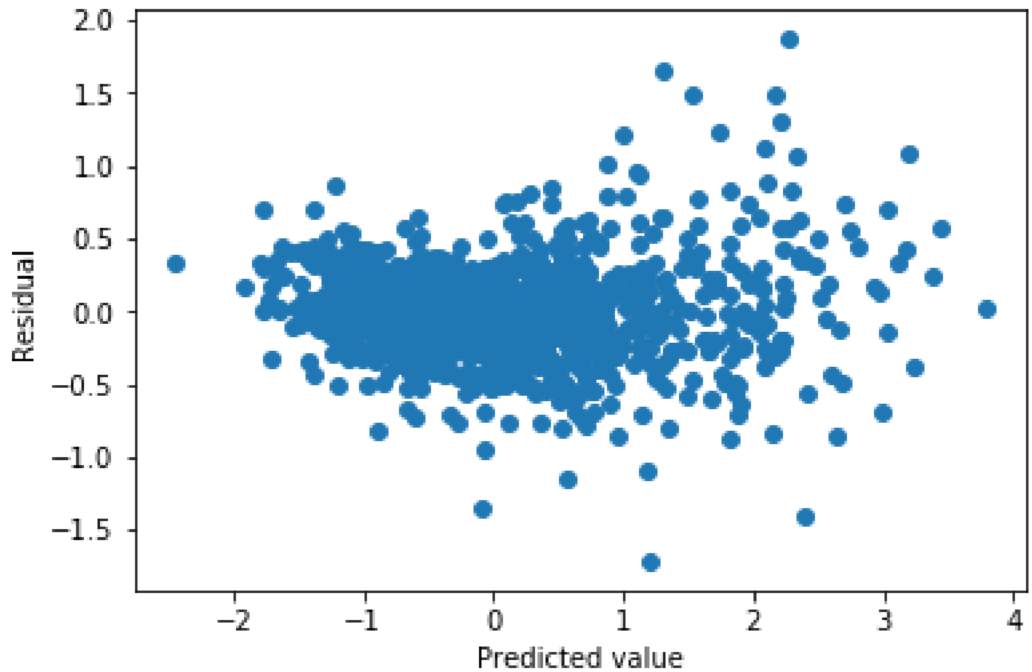
[ ]:
```
import statsmodels.formula.api as smf
linear_results = smf.ols(optimal_linear_model.model.formula, data=data_sig).
↪fit()
```

Now we have the optimal linear model. See Appendix for results details.

3.5 Diagnostics and remedies

In this part, we will do some diagnotics on the linear model. First, we will check the residuals plot.

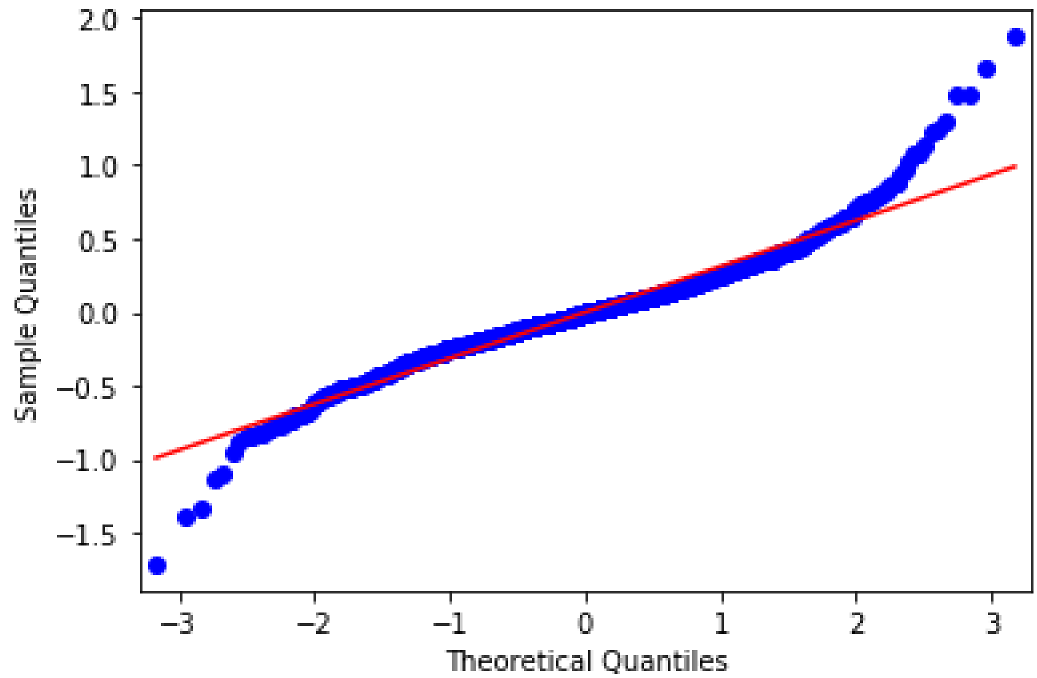
[327]:
```
res = linear_results.resid
plt.scatter(linear_results.fittedvalues,res)
plt.xlabel('Predicted value')
plt.ylabel('Residual');
```



It can be seen that a few samples with higher predicted values have larger residuals. Since most residuals look normal, I won't worry much about these samples.

The following figure shows the QQ plot.

[328]:
```
fig = sm.graphics.qqplot(res, line='s')
plt.show()
```



Notice the points fall along a line in the middle of the graph, but curve off in the extremities. Normal Q-Q plots that exhibit this behavior usually mean the data have more extreme values than would be expected if they truly came from a Normal distribution. Again, sincs most points look normal, we could ignore the extreme cases for now.

4 Interaction terms

In this section, we will add interaction terms into the linear regression model. To simplify the model, we will only include two-way interactions between the numerical variables. We will start with a full model including all regression terms, and then use forward selection method the choose the optimal model.

4.1 Full model with all interaction terms

We list all the numerical variables as follows, and add their interaction terms into the optimal linear model and consider this model as a full model:

[ ]:
```
var = ['OverallQual', 'GarageArea', 'LotArea', 'TotRmsAbvGrd', 'YearRemodAdd',␣
 ↪'Fireplaces', 'FullBath',
        'BsmtFullBath', 'OpenPorchSF', 'MasVnrArea', 'WoodDeckSF',␣
 ↪'OverallCond', 'ScreenPorch']
fomula = optimal_linear_model.model.formula
for i in range(len(var)-1):
    for j in range(i+1,len(var)):
        fomula += '+ ' + var[i] + '*' + var[j]
```

```
inter_results = smf.ols(fomula, data=data_sig).fit()
```

4.2 Model selection using forward selection method

We construct a forward select function for the model with interaction terms. Note that we keep all the linear terms from the optimal linear model we choose in Section 3, and add interaction terms to the optimal linear model.

[332]:
```
def forward_selected_interact(data, response):
    remaining = set()
    for i in range(len(var)-1):
        for j in range(i+1,len(var)):
            remaining.add(var[i] + '*' + var[j])
    selected = []
    current_score, best_new_score = 0.0, 0.0
    while remaining and current_score == best_new_score:

        scores_with_candidates = []
        for candidate in remaining:
            formula = "{} + {} ".format(optimal_linear_model.model.formula, ' +␣
↪'.join(selected + [candidate]))

            model = smf.ols(formula, data)

            res = model.fit()
            score = res.rsquared_adj
            scores_with_candidates.append((score, candidate))

        scores_with_candidates.sort()
        best_new_score, best_candidate = scores_with_candidates.pop()
        if current_score < best_new_score:
            remaining.remove(best_candidate)
            selected.append(best_candidate)
            current_score = best_new_score


    formula = "{} + {} ".format(optimal_linear_model.model.formula,
                                ' + '.join(selected))
    model = smf.ols(formula, data).fit()
    return model

# Find the optimal model
optimal_linear_model_interaction = forward_selected_interact(data_sig,␣
↪'SalePrice')
print(optimal_linear_model_interaction.model.formula)
print(optimal_linear_model_interaction.rsquared_adj)
```

```
SalePrice ~ OverallQual + LotArea + GarageArea + BsmtQual_Ex + TotRmsAbvGrd +
BsmtFinType1_Unf + YearRemodAdd + Fireplaces + Neighborhood_NoRidge +
KitchenQual_Ex + BsmtExposure_Gd + Neighborhood_Crawfor + KitchenAbvGr +
Neighborhood_StoneBr + FullBath + Exterior1st_BrkFace + Neighborhood_NridgHt +
SaleType_WD + BsmtFullBath + OpenPorchSF + MasVnrArea + Functional_Typ +
WoodDeckSF + SaleCondition_Abnorml + KitchenQual_Gd + BldgType_Twnhs +
Foundation_BrkTil + Neighborhood_BrkSide + MSZoning_FV + OverallCond +
Condition1_Norm + MSZoning_Call + GarageType_2Types + ScreenPorch +
LotConfig_CulDSac + GarageType_BuiltIn + HouseStyle_SLvl + BsmtFinType1_GLQ +
Exterior2nd_WdShng + ExterQual_Ex + LotConfig_FR2 + BsmtExposure_No +
HouseStyle_SFoyer + SaleCondition_Alloca + MasVnrType_BrkCmn + LotConfig_FR3 +
Neighborhood_NWAmes + HalfBath + Condition1_RRAe + LandContour_Low +
Exterior1st_WdSdng + Functional_Min2 + Neighborhood_Gilbert + BsmtQual_Gd +
Neighborhood_MeadowV + Exterior2nd_CmentBd + Neighborhood_Mitchel +
LandSlope_Sev + Exterior2nd_WdSdng + Neighborhood_NAmes + Neighborhood_Edwards +
MSZoning_RL + Condition1_PosN + MasVnrType_Stone + Exterior1st_Stone +
Neighborhood_Timber + Exterior1st_CemntBd + GarageFinish_RFn + MSZoning_RH +
EnclosedPorch + 1 + OverallQual*GarageArea + OverallQual*LotArea +
OverallQual*YearRemodAdd + TotRmsAbvGrd*Fireplaces + YearRemodAdd*OverallCond +
Fireplaces*OverallCond + OverallQual*BsmtFullBath + OpenPorchSF*WoodDeckSF +
MasVnrArea*OverallCond + TotRmsAbvGrd*FullBath + LotArea*FullBath +
OverallQual*Fireplaces + MasVnrArea*ScreenPorch + YearRemodAdd*FullBath +
OverallQual*ScreenPorch + Fireplaces*WoodDeckSF + TotRmsAbvGrd*WoodDeckSF +
LotArea*OpenPorchSF + LotArea*Fireplaces + Fireplaces*MasVnrArea +
GarageArea*BsmtFullBath + LotArea*BsmtFullBath + GarageArea*YearRemodAdd +
LotArea*YearRemodAdd + BsmtFullBath*ScreenPorch + YearRemodAdd*WoodDeckSF +
YearRemodAdd*OpenPorchSF + OpenPorchSF*ScreenPorch + TotRmsAbvGrd*ScreenPorch +
OverallCond*ScreenPorch
0.9124256697609201
```

Now we have the optimal model with interaction terms. Next, we will do diagnostics on this model and remedies if needed.

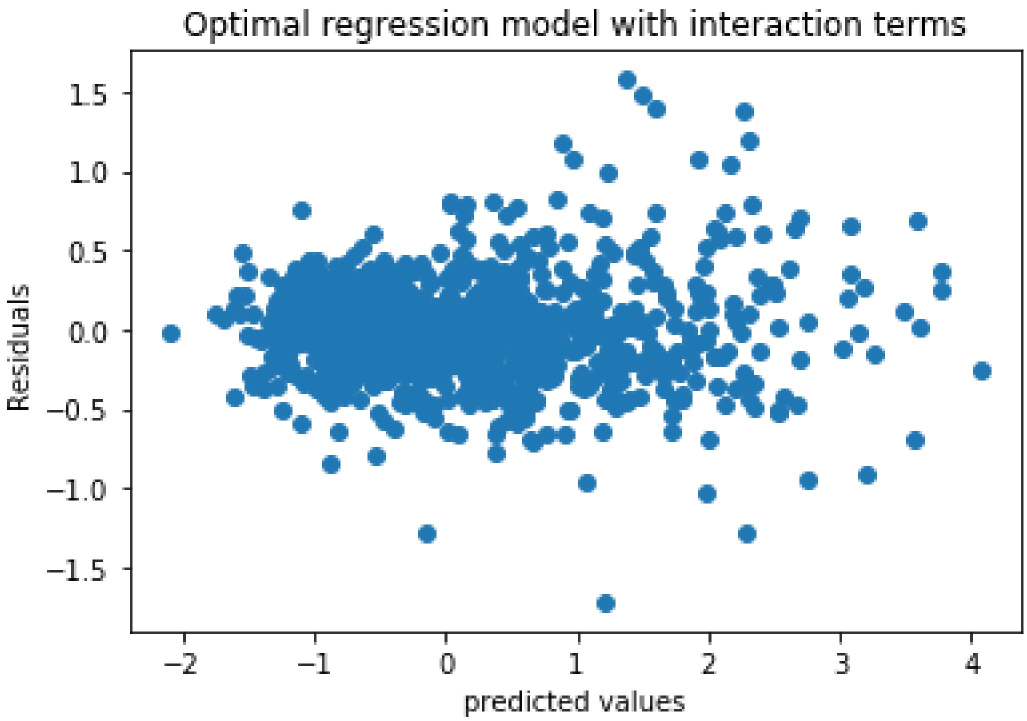
5 Diagnostics and remedies

5.1 Residual plot

We will check the constant variance assumption by plotting the residuals against the predicted sale price.

[339]:
```
res = optimal_linear_model_interaction.resid
plt.scatter(optimal_linear_model_interaction.fittedvalues,res)
plt.xlabel("predicted values")
plt.ylabel("Residuals");
plt.title("Optimal regression model with interaction terms");
```

[339]:
```
Text(0.5, 1.0, 'Optimal regression model with interaction terms')
```

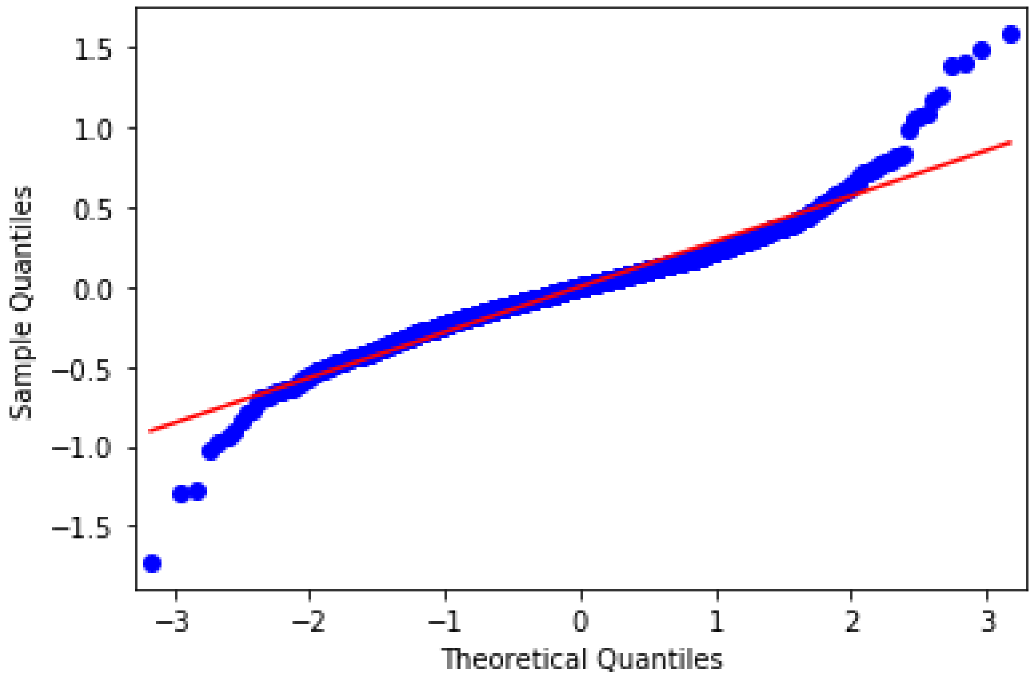


It can be seen that a few samples with higher predicted values have larger residuals. Since most residuals look normal, I won't worry much about these samples.

5.2 QQ plot

To check the assumption on normality, we do a QQ plot.

[340]:
```
fig = sm.graphics.qqplot(res, line='s')
plt.show()
```



Notice the points fall along a line in the middle of the graph, but curve off in the extremities. Normal Q-Q plots that exhibit this behavior usually mean the data have more extreme values than would be expected if they truly came from a Normal distribution.

5.3 Influence statistics

[341]:
```
infl = optimal_linear_model_interaction.get_influence()
df_infl = infl.summary_frame()
df_infl.head()
```

[341]:

	dfb_Intercept	dfb_OverallQual	dfb_LotArea	dfb_GarageArea
0	0.007556	-0.003459	0.003456	0.008259
1	0.003143	0.000073	-0.005052	0.001038
2	0.002968	-0.000028	0.003964	0.000122
3	0.041624	-0.064572	0.030048	-0.113650
4	-0.041402	-0.007497	0.049755	-0.090744

	dfb_BsmtQual_Ex	dfb_TotRmsAbvGrd	dfb_BsmtFinType1_Unf	dfb_YearRemodAdd
0	0.000347	-0.019294	0.001201	-0.009858
1	0.004133	0.001523	-0.002713	-0.007659
2	-0.002094	-0.006559	0.001910	0.001984
3	0.029649	-0.012402	0.083004	0.037642
4	0.133265	-0.100794	-0.029303	-0.023923

```
   dfb_Fireplaces  dfb_Neighborhood_NoRidge  …  \
```

```
0        0.023070                   0.020221  …
1        0.002429                  -0.000618  …
2        0.004877                  -0.006436  …
3        0.034625                   0.024811  …
4       -0.018668                  -0.428452  …

   dfb_YearRemodAdd:OpenPorchSF  dfb_OpenPorchSF:ScreenPorch  \
0                     -0.007501                    -0.006225
1                      0.000920                     0.000328
2                     -0.000579                     0.001288
3                      0.024380                     0.006223
4                     -0.000557                     0.000650

   dfb_TotRmsAbvGrd:ScreenPorch  dfb_OverallCond:ScreenPorch   cooks_d  \
0                      0.000650                     0.001413  0.000082
1                     -0.000141                    -0.004584  0.000017
2                      0.002389                    -0.000152  0.000006
3                     -0.012091                    -0.018493  0.002746
4                      0.035131                     0.020863  0.008360

   standard_resid  hat_diag  dffits_internal  student_resid    dffits
0       -0.603416  0.022109        -0.090732      -0.603258 -0.090708
1        0.147208  0.072849         0.041264       0.147148  0.041247
2        0.176842  0.019826         0.025151       0.176772  0.025141
3       -1.461961  0.114847        -0.526606      -1.462647 -0.526853
4       -3.594706  0.061339        -0.918918      -3.612503 -0.923467

[5 rows x 107 columns]
```

5.4 Identifying Outlying X: Hat Matrix Diagonals

The Hat Matrix Diagonal is also know as the leverage of i-th case. It meansures how much Y_i contributes to the prediction of $\bar{Y}_i$. Observations with large h_{ii} are considered influential, i.e., $h_{ii} > 2p/n$, where p is the number of predictors, and n is number of all observations.

[344]:
```python
# Find the criteria
p = 100
n = len(data_sig)
crit = 2*p/n
print(crit)
```

```
0.1520912547528517
```

[358]:
```python
# Locate the outliers
outliers = data_sig.loc[df_infl['hat_diag'] > crit]
outliers
```

[358]:

	LotArea	OverallQual	OverallCond	YearRemodAdd	MasVnrArea	\
8	-0.841921	0.618742	-0.554698	-1.749664	-0.602333	
29	-0.317137	-1.691892	-1.479430	-1.749664	-0.602333	
54	0.825482	2.929377	-0.554698	1.006214	5.216648	
66	0.818646	0.618742	0.370033	-0.617786	5.690745	
71	-1.839452	-1.691892	-0.554698	-0.617786	-0.602333	
…	…	…	…	…	…	
1192	0.110629	-0.151469	1.294764	-1.749664	-0.602333	
1221	1.499102	2.159165	-0.554698	0.957002	-0.229828	
1247	1.489180	0.618742	-0.554698	-0.371725	0.436166	
1281	2.150231	-0.151469	1.294764	-0.962270	-0.602333	
1312	-0.197627	0.618742	3.144227	1.006214	-0.602333	

	BsmtFullBath	FullBath	HalfBath	KitchenAbvGr	TotRmsAbvGrd	…	\
8	-0.847537	0.800986	-0.796838	5.578925	0.963374	…	
29	-0.847537	-1.052966	-0.796838	-0.170249	-0.334594	…	
54	-0.847537	2.654937	1.194502	-0.170249	2.261341	…	
66	1.117614	0.800986	-0.796838	-0.170249	0.963374	…	
71	1.117614	-1.052966	-0.796838	-0.170249	-0.983577	…	
…	…	…	…	…	…	…	
1192	-0.847537	0.800986	-0.796838	-0.170249	0.963374	…	
1221	1.117614	0.800986	-0.796838	-0.170249	0.963374	…	
1247	1.117614	2.654937	1.194502	-0.170249	3.559308	…	
1281	-0.847537	0.800986	-0.796838	-0.170249	0.963374	…	
1312	-0.847537	0.800986	-0.796838	-0.170249	1.612357	…	

	KitchenQual_Gd	KitchenQual_TA	Functional_Min2	Functional_Typ	\
8	0	1	0	0	
29	0	1	0	1	
54	1	0	0	1	
66	0	1	0	1	
71	0	1	0	1	
…	…	…	…	…	
1192	0	1	0	0	
1221	0	0	0	1	
1247	1	0	0	1	
1281	1	0	0	1	
1312	1	0	0	1	

	GarageType_2Types	GarageType_BuiltIn	GarageFinish_RFn	SaleType_WD	\
8	0	0	0	1	
29	0	0	0	1	
54	0	1	0	0	
66	0	0	0	1	
71	0	1	0	1	
…	…	…	…	…	
1192	0	0	0	1	

```
1221                     0                    0                  0             1
1247                     0                    0                  1             1
1281                     0                    0                  0             1
1312                     0                    0                  1             1

      SaleCondition_Abnorml  SaleCondition_Alloca
8                         1                     0
29                        0                     0
54                        0                     0
66                        0                     0
71                        0                     0
…                       …                     …
1192                      0                     0
1221                      0                     0
1247                      0                     0
1281                      0                     1
1312                      0                     0

[96 rows x 77 columns]
```

The results show that 96 cases are considered outliers.

5.4.1 Identifying Influential Cases using Cook's Distance

[349]:
```python
# F(0.5, p, n-p) = 0.99
# cooks_d
df_infl.loc[df_infl['cooks_d'] > 0.99]['cooks_d']
```

[349]:
```
Series([], Name: cooks_d, dtype: float64)
```

None of the observations appear to have an undue amount of influence.

5.4.2 Multicollinearity Diagnostics: VIF

We use VIF again to check diagnose multicollinearity in the final regression model.

[353]:
```python
exog_ = sm.add_constant(data_sig.drop(['SalePrice'],axis = 1))
vif_ = pd.DataFrame()
vif_["VIF Factor"] = [variance_inflation_factor(exog_.values, i) for i in␣
 ↪range(exog_.shape[1])]
vif_["features"] = exog_.columns
corelated_ = vif_.loc[vif_['VIF Factor']>5]['features'].values
```

[354]:
```python
corelated_
```

[354]:
```
array(['Exterior1st_CemntBd', 'Exterior1st_WdSdng', 'Exterior2nd_CmentBd',
       'Exterior2nd_WdSdng', 'BsmtQual_Ex', 'BsmtQual_Fa', 'BsmtQual_Gd',
       'BsmtQual_TA', 'BsmtExposure_Av', 'BsmtExposure_Gd',
```

```
'BsmtExposure_Mn', 'BsmtExposure_No', 'KitchenQual_Ex',
'KitchenQual_Fa', 'KitchenQual_Gd', 'KitchenQual_TA'], dtype=object)
```

The results turn out that several variables are correlated. However, we can safely ignore this because the variables with high VIFs are dummy variables that represent a categorical variable with three or more categories. If the proportion of cases in the reference category is small, the indicator variables will necessarily have high VIFs, even if the categorical variable is not associated with other variables in the regression model. Therefore, we don't need to do anything about it.

6 Summary

This report analyzed the data on house price and the related variables. Before building the house price predicting model, we clean the original data set and conduct an exploratory data analysis. Then we develop the linear regression model in the following sequence:

- Build a linear regression model with all variables as the full model
- Build a reduced model with less variables
- Propose a null hypothesis that the coefficients of the variables that are not included in the full model are all zero
- Test the hypothesis by F statistics

The test result fails to reject the null hypothesis, so we continue to use the reduced model for further model selection with forward method. After that, we obtain an optimal linear regression model. The coefficients are attached in the Appendix (A.3).

Next, we add interaction terms to the linear model. To simplify the problem, we only consider the two-way interactions. Forward selection mothod is applied to find the best model. The coefficients and confidence intervals are attached in the Appendix (A.5). The adjusted Rsqaured is 0.91.

Our final model is:

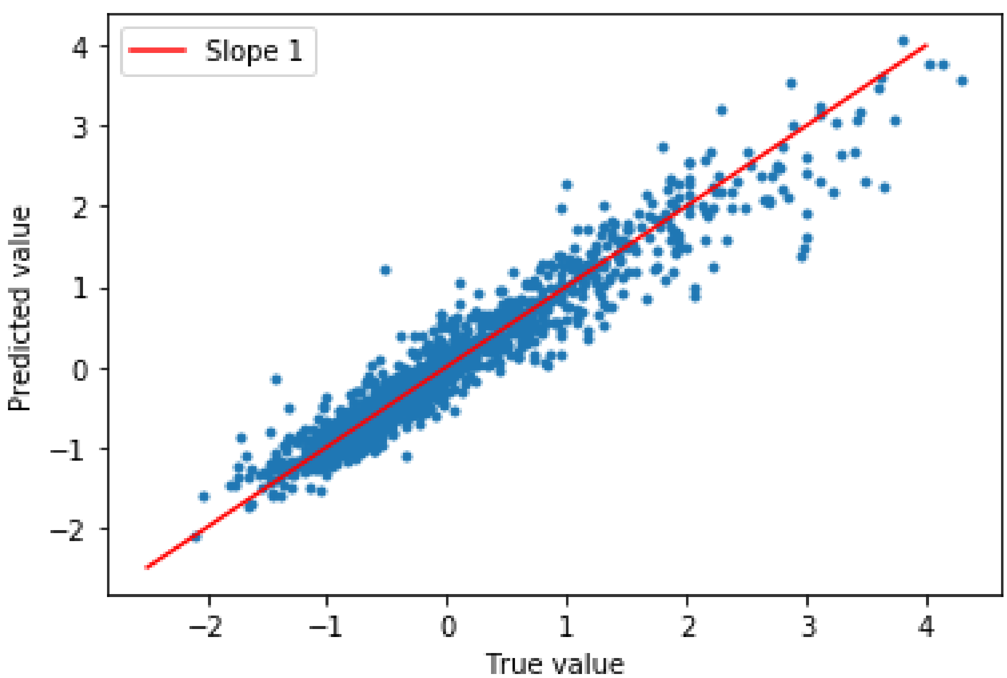
[400]:
```
print(optimal_linear_model_interaction.model.formula)
```

```
SalePrice ~ OverallQual + LotArea + GarageArea + BsmtQual_Ex + TotRmsAbvGrd +
BsmtFinType1_Unf + YearRemodAdd + Fireplaces + Neighborhood_NoRidge +
KitchenQual_Ex + BsmtExposure_Gd + Neighborhood_Crawfor + KitchenAbvGr +
Neighborhood_StoneBr + FullBath + Exterior1st_BrkFace + Neighborhood_NridgHt +
SaleType_WD + BsmtFullBath + OpenPorchSF + MasVnrArea + Functional_Typ +
WoodDeckSF + SaleCondition_Abnorml + KitchenQual_Gd + BldgType_Twnhs +
Foundation_BrkTil + Neighborhood_BrkSide + MSZoning_FV + OverallCond +
Condition1_Norm + MSZoning_Call + GarageType_2Types + ScreenPorch +
LotConfig_CulDSac + GarageType_BuiltIn + HouseStyle_SLvl + BsmtFinType1_GLQ +
Exterior2nd_WdShng + ExterQual_Ex + LotConfig_FR2 + BsmtExposure_No +
HouseStyle_SFoyer + SaleCondition_Alloca + MasVnrType_BrkCmn + LotConfig_FR3 +
Neighborhood_NWAmes + HalfBath + Condition1_RRAe + LandContour_Low +
Exterior1st_WdSdng + Functional_Min2 + Neighborhood_Gilbert + BsmtQual_Gd +
Neighborhood_MeadowV + Exterior2nd_CmentBd + Neighborhood_Mitchel +
LandSlope_Sev + Exterior2nd_WdSdng + Neighborhood_NAmes + Neighborhood_Edwards +
MSZoning_RL + Condition1_PosN + MasVnrType_Stone + Exterior1st_Stone +
Neighborhood_Timber + Exterior1st_CemntBd + GarageFinish_RFn + MSZoning_RH +
```

```
EnclosedPorch + 1 + OverallQual*GarageArea + OverallQual*LotArea +
OverallQual*YearRemodAdd + TotRmsAbvGrd*Fireplaces + YearRemodAdd*OverallCond +
Fireplaces*OverallCond + OverallQual*BsmtFullBath + OpenPorchSF*WoodDeckSF +
MasVnrArea*OverallCond + TotRmsAbvGrd*FullBath + LotArea*FullBath +
OverallQual*Fireplaces + MasVnrArea*ScreenPorch + YearRemodAdd*FullBath +
OverallQual*ScreenPorch + Fireplaces*WoodDeckSF + TotRmsAbvGrd*WoodDeckSF +
LotArea*OpenPorchSF + LotArea*Fireplaces + Fireplaces*MasVnrArea +
GarageArea*BsmtFullBath + LotArea*BsmtFullBath + GarageArea*YearRemodAdd +
LotArea*YearRemodAdd + BsmtFullBath*ScreenPorch + YearRemodAdd*WoodDeckSF +
YearRemodAdd*OpenPorchSF + OpenPorchSF*ScreenPorch + TotRmsAbvGrd*ScreenPorch +
OverallCond*ScreenPorch
```

Here we plot a comparison of predicted values and true values:

[399]:
```
ix = np.linspace(-2.5,4,10)
plt.plot(data_sig['SalePrice'],optimal_linear_model_interaction.fittedvalues, ".
 ↪")
plt.plot(ix,ix,'r',label = "Slope 1")
plt.xlabel('True value')
plt.ylabel('Predicted value');
plt.legend();
```



The plot shows that the dots lie closely to the line with slope 1, which means the model predicts the house prices very well.

We also do some diagnostics on the final model. Residual plot and QQ plot look normal except for some extreme values. Hat Matrix Diagonals identify some outliers while Cook's Distance does not detect any cases that concern us. VIF test finds a few correlated variables. However, since these variables are all dummy variables, we can safely ignore this.

7 Appendix

7.1 Results of the base linear model (the full model)

[365]:
```python
print(base_results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              SalePrice   R-squared:                       0.912
Model:                            OLS   Adj. R-squared:                  0.897
Method:                 Least Squares   F-statistic:                     63.78
Date:                Thu, 07 May 2020   Prob (F-statistic):               0.00
Time:                        21:08:34   Log-Likelihood:                -271.47
No. Observations:                1315   AIC:                             910.9
Df Residuals:                    1131   BIC:                             1864.
Df Model:                         183
Covariance Type:            nonrobust
================================================================================
=========
                            coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
---------
level_0               -9.018e-05      0.005     -0.018      0.986      -0.010
0.010
index                  5.602e-05      0.005      0.012      0.990      -0.009
0.009
MSSubClass               -0.0346      0.050     -0.695      0.487      -0.132
0.063
LotArea                   0.1021      0.015      6.849      0.000       0.073
0.131
OverallQual               0.2139      0.019     11.170      0.000       0.176
0.251
OverallCond               0.0609      0.014      4.501      0.000       0.034
0.087
YearRemodAdd              0.0465      0.017      2.744      0.006       0.013
0.080
MasVnrArea                0.0581      0.015      3.972      0.000       0.029
0.087
BsmtFullBath              0.0634      0.013      4.850      0.000       0.038
0.089
BsmtHalfBath              0.0030      0.010      0.292      0.771      -0.017
0.023
```

FullBath	0.1017	0.017	5.911	0.000	0.068	0.135
HalfBath	0.0487	0.015	3.266	0.001	0.019	0.078
BedroomAbvGr	-0.0116	0.016	-0.727	0.467	-0.043	0.020
KitchenAbvGr	-0.0398	0.017	-2.347	0.019	-0.073	-0.007
TotRmsAbvGrd	0.1630	0.019	8.427	0.000	0.125	0.201
Fireplaces	0.0576	0.012	4.788	0.000	0.034	0.081
GarageYrBlt	-0.0129	0.021	-0.629	0.529	-0.053	0.027
GarageArea	0.1189	0.015	7.930	0.000	0.089	0.148
WoodDeckSF	0.0417	0.011	3.938	0.000	0.021	0.062
OpenPorchSF	0.0385	0.011	3.539	0.000	0.017	0.060
EnclosedPorch	0.0144	0.011	1.313	0.190	-0.007	0.036
ScreenPorch	0.0265	0.010	2.651	0.008	0.007	0.046
MiscVal	-0.0033	0.010	-0.331	0.740	-0.023	0.016
MSZoning_C (all)	-0.4293	0.125	-3.445	0.001	-0.674	-0.185
MSZoning_FV	0.1346	0.085	1.580	0.114	-0.033	0.302
MSZoning_RH	0.1342	0.098	1.364	0.173	-0.059	0.327
MSZoning_RL	0.0622	0.048	1.291	0.197	-0.032	0.157
MSZoning_RM	0.0482	0.052	0.922	0.357	-0.054	0.151
LotShape_IR1	-0.0557	0.039	-1.438	0.151	-0.132	0.020
LotShape_IR2	0.0226	0.055	0.410	0.682	-0.086	0.131
LotShape_IR3	0.0188	0.098	0.192	0.848	-0.173	0.211
LotShape_Reg	-0.0358	0.040	-0.891	0.373	-0.115	0.043
LandContour_Bnk	-0.0038	0.049	-0.076	0.939	-0.101	0.093
LandContour_HLS	0.0653	0.049	1.344	0.179	-0.030	0.161

LandContour_Low	-0.1395	0.066	-2.105	0.036	-0.270	-0.009
LandContour_Lvl	0.0279	0.035	0.791	0.429	-0.041	0.097
LotConfig_Corner	0.0530	0.042	1.269	0.205	-0.029	0.135
LotConfig_CulDSac	0.1697	0.049	3.449	0.001	0.073	0.266
LotConfig_FR2	-0.1136	0.055	-2.068	0.039	-0.221	-0.006
LotConfig_FR3	-0.1981	0.140	-1.414	0.158	-0.473	0.077
LotConfig_Inside	0.0388	0.039	0.999	0.318	-0.037	0.115
LandSlope_Gtl	0.0228	0.060	0.383	0.702	-0.094	0.140
LandSlope_Mod	0.1068	0.059	1.799	0.072	-0.010	0.223
LandSlope_Sev	-0.1798	0.107	-1.677	0.094	-0.390	0.031
Neighborhood_Blmngtn	-0.0126	0.100	-0.125	0.900	-0.209	0.184
Neighborhood_Blueste	-0.0038	0.239	-0.016	0.987	-0.473	0.465
Neighborhood_BrDale	-0.0304	0.112	-0.272	0.786	-0.249	0.189
Neighborhood_BrkSide	0.0922	0.071	1.291	0.197	-0.048	0.232
Neighborhood_ClearCr	-0.0449	0.086	-0.521	0.603	-0.214	0.124
Neighborhood_CollgCr	-0.1124	0.043	-2.625	0.009	-0.196	-0.028
Neighborhood_Crawfor	0.2814	0.059	4.739	0.000	0.165	0.398
Neighborhood_Edwards	-0.1516	0.049	-3.079	0.002	-0.248	-0.055
Neighborhood_Gilbert	-0.1670	0.055	-3.038	0.002	-0.275	-0.059
Neighborhood_IDOTRR	-0.1017	0.097	-1.050	0.294	-0.292	0.088
Neighborhood_MeadowV	-0.1898	0.123	-1.538	0.124	-0.432	0.052
Neighborhood_Mitchel	-0.1553	0.059	-2.620	0.009	-0.272	-0.039
Neighborhood_NAmes	-0.1726	0.039	-4.434	0.000	-0.249	-0.096
Neighborhood_NPkVill	0.0352	0.172	0.205	0.837	-0.302	0.372

Neighborhood_NWAmes	-0.1709	0.050	-3.428	0.001	-0.269	-0.073
Neighborhood_NoRidge	0.5267	0.068	7.719	0.000	0.393	0.661
Neighborhood_NridgHt	0.2140	0.061	3.510	0.000	0.094	0.334
Neighborhood_OldTown	-0.1201	0.069	-1.737	0.083	-0.256	0.016
Neighborhood_SWISU	-0.1302	0.092	-1.412	0.158	-0.311	0.051
Neighborhood_Sawyer	-0.0688	0.051	-1.356	0.175	-0.168	0.031
Neighborhood_SawyerW	0.0195	0.055	0.354	0.723	-0.088	0.127
Neighborhood_Somerst	-0.0115	0.086	-0.134	0.893	-0.179	0.156
Neighborhood_StoneBr	0.5352	0.083	6.428	0.000	0.372	0.699
Neighborhood_Timber	-0.1652	0.066	-2.495	0.013	-0.295	-0.035
Neighborhood_Veenker	0.0544	0.113	0.481	0.631	-0.168	0.276
Condition1_Artery	-0.0028	0.067	-0.041	0.967	-0.135	0.130
Condition1_Feedr	0.0141	0.057	0.245	0.806	-0.098	0.126
Condition1_Norm	0.1102	0.042	2.639	0.008	0.028	0.192
Condition1_PosA	-0.0097	0.123	-0.079	0.937	-0.251	0.231
Condition1_PosN	0.1665	0.083	2.015	0.044	0.004	0.329
Condition1_RRAe	-0.2872	0.113	-2.551	0.011	-0.508	-0.066
Condition1_RRAn	0.0119	0.074	0.160	0.873	-0.134	0.158
Condition1_RRNe	-0.0910	0.214	-0.425	0.671	-0.511	0.329
Condition1_RRNn	0.0379	0.155	0.244	0.807	-0.266	0.342
BldgType_1Fam	0.0824	0.097	0.854	0.393	-0.107	0.272
BldgType_2fmCon	0.0450	0.109	0.412	0.680	-0.170	0.260
BldgType_Duplex	0.0103	0.093	0.111	0.912	-0.172	0.192
BldgType_Twnhs	-0.1230	0.077	-1.606	0.109	-0.273	0.027

BldgType_TwnhsE	-0.0648	0.061	-1.070	0.285	-0.184	0.054
HouseStyle_1.5Fin	0.0050	0.045	0.112	0.911	-0.083	0.093
HouseStyle_1.5Unf	-0.0613	0.111	-0.553	0.580	-0.279	0.156
HouseStyle_1Story	0.0174	0.056	0.312	0.755	-0.092	0.127
HouseStyle_2.5Fin	0.1753	0.140	1.257	0.209	-0.098	0.449
HouseStyle_2.5Unf	0.0993	0.118	0.841	0.400	-0.132	0.331
HouseStyle_2Story	-0.0401	0.040	-1.012	0.312	-0.118	0.038
HouseStyle_SFoyer	-0.1524	0.078	-1.945	0.052	-0.306	0.001
HouseStyle_SLvl	-0.0933	0.060	-1.550	0.122	-0.212	0.025
Exterior1st_AsbShng	0.0857	0.173	0.494	0.621	-0.254	0.426
Exterior1st_BrkComm	-0.3905	0.407	-0.961	0.337	-1.188	0.407
Exterior1st_BrkFace	0.3373	0.088	3.835	0.000	0.165	0.510
Exterior1st_CBlock	0.1075	0.184	0.584	0.560	-0.254	0.469
Exterior1st_CemntBd	-0.4316	0.240	-1.796	0.073	-0.903	0.040
Exterior1st_HdBoard	-0.0527	0.079	-0.667	0.505	-0.208	0.102
Exterior1st_ImStucc	-0.1689	0.329	-0.513	0.608	-0.815	0.477
Exterior1st_MetalSd	0.0103	0.132	0.078	0.938	-0.249	0.269
Exterior1st_Plywood	0.0480	0.081	0.592	0.554	-0.111	0.207
Exterior1st_Stone	0.4777	0.269	1.773	0.076	-0.051	1.006
Exterior1st_Stucco	-0.0060	0.128	-0.047	0.963	-0.257	0.245
Exterior1st_VinylSd	0.0703	0.111	0.633	0.527	-0.148	0.288
Exterior1st_Wd Sdng	-0.1263	0.077	-1.647	0.100	-0.277	0.024
Exterior1st_WdShing	-0.0108	0.106	-0.102	0.919	-0.218	0.197
Exterior2nd_AsbShng	-0.0366	0.164	-0.224	0.823	-0.358	0.284

Exterior2nd_AsphShn	0.1183	0.235	0.503	0.615	-0.343
0.580					
Exterior2nd_Brk Cmn	-0.0506	0.222	-0.228	0.820	-0.486
0.385					
Exterior2nd_BrkFace	0.0028	0.108	0.026	0.980	-0.210
0.216					
Exterior2nd_CBlock	0.1075	0.184	0.584	0.560	-0.254
0.469					
Exterior2nd_CmentBd	0.5325	0.243	2.189	0.029	0.055
1.010					
Exterior2nd_HdBoard	-0.0252	0.074	-0.342	0.733	-0.170
0.119					
Exterior2nd_ImStucc	0.0037	0.128	0.029	0.977	-0.247
0.254					
Exterior2nd_MetalSd	0.0112	0.127	0.088	0.930	-0.238
0.261					
Exterior2nd_Other	-0.2707	0.333	-0.814	0.416	-0.923
0.382					
Exterior2nd_Plywood	-0.0737	0.070	-1.049	0.294	-0.212
0.064					
Exterior2nd_Stone	-0.2433	0.268	-0.907	0.365	-0.770
0.283					
Exterior2nd_Stucco	-0.0054	0.125	-0.043	0.966	-0.250
0.239					
Exterior2nd_VinylSd	-0.0609	0.102	-0.599	0.549	-0.261
0.139					
Exterior2nd_Wd Sdng	0.0922	0.072	1.288	0.198	-0.048
0.233					
Exterior2nd_Wd Shng	-0.1519	0.086	-1.760	0.079	-0.321
0.017					
MasVnrType_BrkCmn	-0.1560	0.075	-2.072	0.039	-0.304
-0.008					
MasVnrType_BrkFace	0.0268	0.032	0.845	0.398	-0.035
0.089					
MasVnrType_None	0.0264	0.033	0.801	0.423	-0.038
0.091					
MasVnrType_Stone	0.0527	0.039	1.351	0.177	-0.024
0.129					
ExterQual_Ex	0.1634	0.070	2.318	0.021	0.025
0.302					
ExterQual_Fa	-0.1740	0.134	-1.300	0.194	-0.437
0.089					
ExterQual_Gd	0.0182	0.051	0.356	0.722	-0.082
0.119					
ExterQual_TA	-0.0578	0.051	-1.129	0.259	-0.158
0.043					
ExterCond_Ex	0.0091	0.193	0.047	0.963	-0.370
0.388					

ExterCond_Fa	0.0469	0.101	0.464	0.643	-0.151	0.245
ExterCond_Gd	-0.0848	0.070	-1.218	0.223	-0.221	0.052
ExterCond_TA	-0.0213	0.067	-0.316	0.752	-0.153	0.111
Foundation_BrkTil	-0.1599	0.058	-2.748	0.006	-0.274	-0.046
Foundation_CBlock	0.0050	0.055	0.090	0.928	-0.104	0.114
Foundation_PConc	0.0141	0.055	0.256	0.798	-0.094	0.123
Foundation_Stone	0.1476	0.129	1.142	0.254	-0.106	0.401
Foundation_Wood	-0.0569	0.164	-0.347	0.729	-0.379	0.265
BsmtQual_Ex	0.2076	0.044	4.723	0.000	0.121	0.294
BsmtQual_Fa	-0.1057	0.058	-1.834	0.067	-0.219	0.007
BsmtQual_Gd	-0.0617	0.029	-2.121	0.034	-0.119	-0.005
BsmtQual_TA	-0.0902	0.031	-2.946	0.003	-0.150	-0.030
BsmtCond_Fa	-0.0306	0.067	-0.458	0.647	-0.162	0.101
BsmtCond_Gd	-0.0445	0.065	-0.689	0.491	-0.171	0.082
BsmtCond_Po	0.0232	0.163	0.143	0.887	-0.296	0.342
BsmtCond_TA	0.0018	0.057	0.031	0.975	-0.110	0.114
BsmtExposure_Av	-0.0563	0.024	-2.299	0.022	-0.104	-0.008
BsmtExposure_Gd	0.1901	0.032	5.934	0.000	0.127	0.253
BsmtExposure_Mn	-0.0632	0.030	-2.087	0.037	-0.123	-0.004
BsmtExposure_No	-0.1207	0.022	-5.616	0.000	-0.163	-0.079
BsmtFinType1_ALQ	-0.0076	0.026	-0.288	0.773	-0.059	0.044
BsmtFinType1_BLQ	-0.0043	0.030	-0.145	0.885	-0.062	0.054
BsmtFinType1_GLQ	0.0756	0.027	2.769	0.006	0.022	0.129
BsmtFinType1_LwQ	-0.0184	0.039	-0.467	0.641	-0.096	0.059

BsmtFinType1_Rec	0.0015	0.031	0.048	0.961	-0.059	0.062
BsmtFinType1_Unf	-0.0969	0.025	-3.876	0.000	-0.146	-0.048
BsmtFinType2_ALQ	-0.0202	0.074	-0.273	0.785	-0.165	0.125
BsmtFinType2_BLQ	-0.0317	0.058	-0.546	0.585	-0.146	0.082
BsmtFinType2_GLQ	-0.0312	0.094	-0.333	0.739	-0.215	0.153
BsmtFinType2_LwQ	0.0065	0.052	0.124	0.901	-0.095	0.108
BsmtFinType2_Rec	0.0046	0.050	0.092	0.927	-0.094	0.103
BsmtFinType2_Unf	0.0219	0.032	0.682	0.496	-0.041	0.085
Heating_GasA	-0.0127	0.116	-0.109	0.913	-0.241	0.216
Heating_GasW	0.1126	0.128	0.881	0.378	-0.138	0.363
Heating_Grav	0.0641	0.194	0.331	0.741	-0.316	0.444
Heating_OthW	-0.2141	0.286	-0.748	0.455	-0.776	0.347
HeatingQC_Ex	0.0043	0.078	0.055	0.956	-0.149	0.157
HeatingQC_Fa	-0.0527	0.092	-0.575	0.566	-0.233	0.127
HeatingQC_Gd	-0.0477	0.079	-0.607	0.544	-0.202	0.107
HeatingQC_Po	0.0902	0.311	0.290	0.772	-0.520	0.700
HeatingQC_TA	-0.0442	0.077	-0.571	0.568	-0.196	0.108
CentralAir_N	-0.0690	0.035	-1.947	0.052	-0.139	0.001
CentralAir_Y	0.0189	0.040	0.478	0.633	-0.059	0.096
Electrical_FuseA	0.0278	0.085	0.328	0.743	-0.139	0.194
Electrical_FuseF	0.0013	0.102	0.013	0.990	-0.198	0.201
Electrical_FuseP	-0.1180	0.219	-0.538	0.591	-0.548	0.312
Electrical_Mix	0.0232	0.163	0.143	0.887	-0.296	0.342
Electrical_SBrkr	0.0156	0.082	0.192	0.848	-0.144	0.176

KitchenQual_Ex	0.2275	0.045	5.030	0.000	0.139	0.316
KitchenQual_Fa	-0.0829	0.062	-1.336	0.182	-0.205	0.039
KitchenQual_Gd	-0.0872	0.030	-2.897	0.004	-0.146	-0.028
KitchenQual_TA	-0.1075	0.029	-3.727	0.000	-0.164	-0.051
Functional_Maj1	0.0550	0.117	0.469	0.639	-0.175	0.285
Functional_Maj2	-0.0837	0.186	-0.450	0.653	-0.449	0.282
Functional_Min1	0.0258	0.088	0.293	0.769	-0.147	0.198
Functional_Min2	0.1273	0.089	1.433	0.152	-0.047	0.302
Functional_Mod	0.0797	0.121	0.659	0.510	-0.158	0.317
Functional_Sev	-0.4143	0.332	-1.248	0.212	-1.066	0.237
Functional_Typ	0.1601	0.069	2.310	0.021	0.024	0.296
GarageType_2Types	-0.3650	0.133	-2.748	0.006	-0.626	-0.104
GarageType_Attchd	0.0598	0.044	1.367	0.172	-0.026	0.146
GarageType_Basment	-0.0267	0.082	-0.328	0.743	-0.187	0.133
GarageType_BuiltIn	0.1733	0.056	3.093	0.002	0.063	0.283
GarageType_CarPort	0.0574	0.135	0.426	0.670	-0.207	0.322
GarageType_Detchd	0.0511	0.045	1.133	0.257	-0.037	0.140
GarageFinish_Fin	-0.0080	0.023	-0.355	0.723	-0.053	0.036
GarageFinish_RFn	-0.0378	0.021	-1.799	0.072	-0.079	0.003
GarageFinish_Unf	-0.0043	0.023	-0.183	0.855	-0.050	0.041
PavedDrive_N	-0.0334	0.045	-0.746	0.456	-0.121	0.055
PavedDrive_P	-0.0660	0.051	-1.285	0.199	-0.167	0.035
PavedDrive_Y	0.0493	0.035	1.403	0.161	-0.020	0.118
SaleType_COD	-0.1419	0.083	-1.715	0.087	-0.304	0.020

```
SaleType_CWD               0.1470      0.164      0.895      0.371      -0.175
0.469
SaleType_Con               0.1128      0.223      0.507      0.612      -0.324
0.549
SaleType_ConLD            -0.0285      0.145     -0.197      0.844      -0.313
0.256
SaleType_ConLI            -0.1826      0.165     -1.107      0.268      -0.506
0.141
SaleType_ConLw            -0.2937      0.167     -1.763      0.078      -0.621
0.033
SaleType_New               0.2534      0.190      1.337      0.181      -0.118
0.625
SaleType_Oth               0.2857      0.298      0.959      0.338      -0.299
0.870
SaleType_WD               -0.2022      0.065     -3.117      0.002      -0.330
-0.075
SaleCondition_Abnorml     -0.1115      0.080     -1.402      0.161      -0.268
0.044
SaleCondition_AdjLand      0.2489      0.304      0.818      0.413      -0.348
0.845
SaleCondition_Alloca       0.2275      0.138      1.649      0.099      -0.043
0.498
SaleCondition_Family      -0.1184      0.096     -1.228      0.220      -0.307
0.071
SaleCondition_Normal      -0.0196      0.073     -0.268      0.789      -0.163
0.124
SaleCondition_Partial     -0.2769      0.183     -1.513      0.131      -0.636
0.082
==============================================================================
Omnibus:                      148.816   Durbin-Watson:                   1.982
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1006.064
Skew:                           0.268   Prob(JB):                    3.44e-219
Kurtosis:                       7.251   Cond. No.                     1.35e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 9.36e-24. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

7.2 Results of the reduced linear model

[382]:
```python
print(reduced_results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              SalePrice   R-squared:                       0.903
```

```
Model:                            OLS   Adj. R-squared:                  0.897
Method:                 Least Squares   F-statistic:                     157.5
Date:                Thu, 07 May 2020   Prob (F-statistic):               0.00
Time:                        21:15:30   Log-Likelihood:                -335.61
No. Observations:                1315   AIC:                             819.2
Df Residuals:                    1241   BIC:                             1203.
Df Model:                          73
Covariance Type:            nonrobust
================================================================================
=========
                            coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
---------
const                     0.0177      0.041      0.429      0.668      -0.063
0.099
LotArea                   0.1159      0.012      9.352      0.000       0.092
0.140
OverallQual               0.2252      0.017     13.114      0.000       0.192
0.259
OverallCond               0.0553      0.011      4.897      0.000       0.033
0.077
YearRemodAdd              0.0628      0.015      4.245      0.000       0.034
0.092
MasVnrArea                0.0550      0.011      4.869      0.000       0.033
0.077
BsmtFullBath              0.0622      0.012      5.339      0.000       0.039
0.085
FullBath                  0.0915      0.015      6.006      0.000       0.062
0.121
HalfBath                  0.0271      0.011      2.355      0.019       0.005
0.050
KitchenAbvGr             -0.0613      0.011     -5.804      0.000      -0.082
-0.041
TotRmsAbvGrd              0.1629      0.014     11.521      0.000       0.135
0.191
Fireplaces                0.0597      0.011      5.409      0.000       0.038
0.081
GarageArea                0.1243      0.013      9.946      0.000       0.100
0.149
WoodDeckSF                0.0394      0.010      3.910      0.000       0.020
0.059
OpenPorchSF               0.0381      0.010      3.745      0.000       0.018
0.058
EnclosedPorch             0.0113      0.010      1.133      0.258      -0.008
0.031
ScreenPorch               0.0271      0.010      2.837      0.005       0.008
0.046
```

MSZoning_C (all)	-0.4416	0.124	-3.564	0.000	-0.685	-0.199
MSZoning_FV	0.1958	0.056	3.498	0.000	0.086	0.306
MSZoning_RH	0.1162	0.104	1.116	0.265	-0.088	0.321
MSZoning_RL	0.0850	0.035	2.461	0.014	0.017	0.153
LandContour_Low	-0.1425	0.076	-1.866	0.062	-0.292	0.007
LotConfig_CulDSac	0.1096	0.039	2.794	0.005	0.033	0.187
LotConfig_FR2	-0.1458	0.052	-2.818	0.005	-0.247	-0.044
LotConfig_FR3	-0.2612	0.166	-1.575	0.115	-0.587	0.064
LandSlope_Sev	-0.2346	0.138	-1.701	0.089	-0.505	0.036
Neighborhood_BrkSide	0.2101	0.054	3.861	0.000	0.103	0.317
Neighborhood_Crawfor	0.3366	0.054	6.247	0.000	0.231	0.442
Neighborhood_Edwards	-0.1043	0.045	-2.336	0.020	-0.192	-0.017
Neighborhood_Gilbert	-0.1082	0.046	-2.347	0.019	-0.199	-0.018
Neighborhood_MeadowV	-0.2484	0.117	-2.117	0.034	-0.479	-0.018
Neighborhood_Mitchel	-0.1251	0.056	-2.243	0.025	-0.234	-0.016
Neighborhood_NAmes	-0.0887	0.033	-2.649	0.008	-0.154	-0.023
Neighborhood_NWAmes	-0.1599	0.045	-3.530	0.000	-0.249	-0.071
Neighborhood_NoRidge	0.5983	0.062	9.648	0.000	0.477	0.720
Neighborhood_NridgHt	0.2334	0.054	4.354	0.000	0.128	0.339
Neighborhood_StoneBr	0.4807	0.075	6.375	0.000	0.333	0.629
Neighborhood_Timber	-0.0930	0.062	-1.505	0.133	-0.214	0.028
Condition1_Norm	0.0915	0.030	3.005	0.003	0.032	0.151
Condition1_PosN	0.1212	0.084	1.442	0.150	-0.044	0.286
Condition1_RRAe	-0.2727	0.108	-2.516	0.012	-0.485	-0.060

BldgType_Twnhs	-0.2062	0.060	-3.431	0.001	-0.324	-0.088
HouseStyle_SFoyer	-0.2308	0.068	-3.381	0.001	-0.365	-0.097
HouseStyle_SLvl	-0.1563	0.046	-3.375	0.001	-0.247	-0.065
Exterior1st_BrkFace	0.3723	0.056	6.676	0.000	0.263	0.482
Exterior1st_CemntBd	-0.3551	0.239	-1.486	0.138	-0.824	0.114
Exterior1st_Stone	0.3159	0.239	1.324	0.186	-0.152	0.784
Exterior1st_Wd Sdng	-0.1511	0.061	-2.488	0.013	-0.270	-0.032
Exterior2nd_CmentBd	0.4855	0.241	2.012	0.044	0.012	0.959
Exterior2nd_Wd Sdng	0.1151	0.061	1.895	0.058	-0.004	0.234
Exterior2nd_Wd Shng	-0.1443	0.064	-2.256	0.024	-0.270	-0.019
MasVnrType_BrkCmn	-0.2198	0.091	-2.423	0.016	-0.398	-0.042
MasVnrType_Stone	0.0528	0.036	1.461	0.144	-0.018	0.124
ExterQual_Ex	0.1517	0.067	2.266	0.024	0.020	0.283
Foundation_BrkTil	-0.1738	0.038	-4.539	0.000	-0.249	-0.099
BsmtQual_Ex	0.2843	0.041	7.013	0.000	0.205	0.364
BsmtQual_Fa	-0.1286	0.051	-2.507	0.012	-0.229	-0.028
BsmtQual_Gd	-0.0487	0.025	-1.925	0.054	-0.098	0.001
BsmtQual_TA	-0.0892	0.026	-3.426	0.001	-0.140	-0.038
BsmtExposure_Av	-0.0356	0.024	-1.469	0.142	-0.083	0.012
BsmtExposure_Gd	0.2103	0.030	6.923	0.000	0.151	0.270
BsmtExposure_Mn	-0.0539	0.029	-1.889	0.059	-0.110	0.002
BsmtExposure_No	-0.1031	0.019	-5.339	0.000	-0.141	-0.065
BsmtFinType1_GLQ	0.0840	0.030	2.846	0.005	0.026	0.142
BsmtFinType1_Unf	-0.0855	0.028	-3.088	0.002	-0.140	-0.031

```
KitchenQual_Ex           0.2629      0.043      6.074      0.000       0.178
0.348
KitchenQual_Fa          -0.0994      0.059     -1.683      0.093      -0.215
0.016
KitchenQual_Gd          -0.0473      0.026     -1.826      0.068      -0.098
0.004
KitchenQual_TA          -0.0985      0.024     -4.021      0.000      -0.147
-0.050
Functional_Min2          0.1179      0.078      1.512      0.131      -0.035
0.271
Functional_Typ           0.1475      0.049      2.990      0.003       0.051
0.244
GarageType_2Types       -0.4237      0.142     -2.975      0.003      -0.703
-0.144
GarageType_BuiltIn       0.1344      0.042      3.207      0.001       0.052
0.217
GarageFinish_RFn        -0.0256      0.022     -1.180      0.238      -0.068
0.017
SaleType_WD             -0.1615      0.030     -5.447      0.000      -0.220
-0.103
SaleCondition_Abnorml   -0.1085      0.039     -2.769      0.006      -0.185
-0.032
SaleCondition_Alloca     0.2830      0.132      2.149      0.032       0.025
0.541
==============================================================================
Omnibus:                      165.037   Durbin-Watson:                   1.933
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1017.508
Skew:                           0.394   Prob(JB):                    1.12e-221
Kurtosis:                       7.237   Cond. No.                     1.11e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 6.22e-29. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

7.3 Results of the optimal linear model

[386]:
```python
# Results of the optimal linear model
print(linear_results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              SalePrice   R-squared:                       0.903
Model:                            OLS   Adj. R-squared:                  0.897
Method:                 Least Squares   F-statistic:                     164.6
Date:                Thu, 07 May 2020   Prob (F-statistic):               0.00
```

```
Time:                        21:22:40   Log-Likelihood:                -335.94
No. Observations:                1315   AIC:                             813.9
Df Residuals:                    1244   BIC:                             1182.
Df Model:                          70
Covariance Type:            nonrobust
================================================================================
=========
                            coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
---------
Intercept                -0.2183      0.073     -3.007      0.003      -0.361
-0.076
OverallQual               0.2264      0.017     13.254      0.000       0.193
0.260
LotArea                   0.1162      0.012      9.391      0.000       0.092
0.140
GarageArea                0.1239      0.012      9.932      0.000       0.099
0.148
BsmtQual_Ex               0.3754      0.053      7.019      0.000       0.270
0.480
TotRmsAbvGrd              0.1633      0.014     11.581      0.000       0.136
0.191
BsmtFinType1_Unf         -0.0882      0.027     -3.224      0.001      -0.142
-0.035
YearRemodAdd              0.0636      0.015      4.346      0.000       0.035
0.092
Fireplaces                0.0594      0.011      5.397      0.000       0.038
0.081
Neighborhood_NoRidge      0.5979      0.062      9.655      0.000       0.476
0.719
KitchenQual_Ex            0.3609      0.053      6.746      0.000       0.256
0.466
BsmtExposure_Gd           0.2512      0.038      6.558      0.000       0.176
0.326
Neighborhood_Crawfor      0.3352      0.054      6.236      0.000       0.230
0.441
KitchenAbvGr             -0.0614      0.011     -5.819      0.000      -0.082
-0.041
Neighborhood_StoneBr      0.4801      0.075      6.373      0.000       0.332
0.628
FullBath                  0.0906      0.015      5.974      0.000       0.061
0.120
Exterior1st_BrkFace       0.3741      0.056      6.730      0.000       0.265
0.483
Neighborhood_NridgHt      0.2326      0.054      4.346      0.000       0.128
0.338
SaleType_WD              -0.1621      0.030     -5.473      0.000      -0.220
```

						-0.104
BsmtFullBath	0.0622	0.012	5.337	0.000	0.039	0.085
OpenPorchSF	0.0383	0.010	3.774	0.000	0.018	0.058
MasVnrArea	0.0551	0.011	4.886	0.000	0.033	0.077
Functional_Typ	0.1503	0.049	3.058	0.002	0.054	0.247
WoodDeckSF	0.0397	0.010	3.942	0.000	0.020	0.059
SaleCondition_Abnorml	-0.1100	0.039	-2.823	0.005	-0.187	-0.034
KitchenQual_Gd	0.0518	0.027	1.889	0.059	-0.002	0.106
BldgType_Twnhs	-0.2056	0.060	-3.425	0.001	-0.323	-0.088
Foundation_BrkTil	-0.1767	0.038	-4.648	0.000	-0.251	-0.102
Neighborhood_BrkSide	0.2122	0.054	3.925	0.000	0.106	0.318
MSZoning_FV	0.1951	0.056	3.491	0.000	0.085	0.305
OverallCond	0.0547	0.011	4.863	0.000	0.033	0.077
Condition1_Norm	0.0934	0.030	3.080	0.002	0.034	0.153
MSZoning_Call	-0.4368	0.123	-3.546	0.000	-0.678	-0.195
GarageType_2Types	-0.4184	0.142	-2.944	0.003	-0.697	-0.140
ScreenPorch	0.0271	0.010	2.842	0.005	0.008	0.046
LotConfig_CulDSac	0.1089	0.039	2.780	0.006	0.032	0.186
GarageType_BuiltIn	0.1345	0.042	3.213	0.001	0.052	0.217
HouseStyle_SLvl	-0.1512	0.045	-3.333	0.001	-0.240	-0.062
BsmtFinType1_GLQ	0.0825	0.029	2.805	0.005	0.025	0.140
Exterior2nd_WdShng	-0.1427	0.064	-2.234	0.026	-0.268	-0.017
ExterQual_Ex	0.1528	0.067	2.285	0.022	0.022	0.284
LotConfig_FR2	-0.1450	0.052	-2.807	0.005	-0.246	-0.044
BsmtExposure_No	-0.0594	0.024	-2.485	0.013	-0.106	

						-0.012
HouseStyle_SFoyer	-0.2237	0.067	-3.333	0.001	-0.355	-0.092
SaleCondition_Alloca	0.2784	0.131	2.125	0.034	0.021	0.535
MasVnrType_BrkCmn	-0.2212	0.091	-2.443	0.015	-0.399	-0.044
LotConfig_FR3	-0.2683	0.165	-1.625	0.104	-0.592	0.056
Neighborhood_NWAmes	-0.1594	0.045	-3.521	0.000	-0.248	-0.071
HalfBath	0.0267	0.011	2.336	0.020	0.004	0.049
Condition1_RRAe	-0.2690	0.108	-2.487	0.013	-0.481	-0.057
LandContour_Low	-0.1404	0.076	-1.846	0.065	-0.290	0.009
Exterior1st_WdSdng	-0.1514	0.061	-2.496	0.013	-0.270	-0.032
Functional_Min2	0.1194	0.078	1.535	0.125	-0.033	0.272
Neighborhood_Gilbert	-0.1072	0.046	-2.330	0.020	-0.198	-0.017
BsmtQual_Gd	0.0420	0.031	1.374	0.170	-0.018	0.102
Neighborhood_MeadowV	-0.2471	0.117	-2.108	0.035	-0.477	-0.017
Exterior2nd_CmentBd	0.4838	0.241	2.006	0.045	0.011	0.957
Neighborhood_Mitchel	-0.1239	0.056	-2.226	0.026	-0.233	-0.015
LandSlope_Sev	-0.2319	0.137	-1.689	0.091	-0.501	0.037
Exterior2nd_WdSdng	0.1154	0.061	1.903	0.057	-0.004	0.234
Neighborhood_NAmes	-0.0881	0.033	-2.638	0.008	-0.154	-0.023
Neighborhood_Edwards	-0.1050	0.044	-2.362	0.018	-0.192	-0.018
MSZoning_RL	0.0852	0.034	2.474	0.014	0.018	0.153
Condition1_PosN	0.1214	0.084	1.445	0.149	-0.043	0.286
MasVnrType_Stone	0.0531	0.036	1.470	0.142	-0.018	0.124
Exterior1st_Stone	0.3220	0.238	1.352	0.177	-0.145	0.789
Neighborhood_Timber	-0.0928	0.062	-1.502	0.133	-0.214	

```
0.028
Exterior1st_CemntBd      -0.3541      0.239     -1.483      0.138      -0.822
0.114
GarageFinish_RFn         -0.0251      0.022     -1.161      0.246      -0.067
0.017
MSZoning_RH               0.1159      0.104      1.115      0.265      -0.088
0.320
EnclosedPorch             0.0109      0.010      1.093      0.275      -0.009
0.030
==============================================================================
Omnibus:                      165.637   Durbin-Watson:                   1.936
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1013.283
Skew:                           0.399   Prob(JB):                    9.30e-221
Kurtosis:                       7.226   Cond. No.                         88.5
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

7.4 full model with all interaction terms

[387]:
```
print(inter_results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              SalePrice   R-squared:                       0.920
Model:                            OLS   Adj. R-squared:                  0.910
Method:                 Least Squares   F-statistic:                     90.65
Date:                Thu, 07 May 2020   Prob (F-statistic):               0.00
Time:                        21:25:23   Log-Likelihood:                -205.89
No. Observations:                1315   AIC:                             709.8
Df Residuals:                    1166   BIC:                             1482.
Df Model:                         148
Covariance Type:            nonrobust
================================================================================
=============
                                coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
-------------
Intercept                    -0.3364      0.072     -4.642      0.000
-0.479      -0.194
OverallQual                   0.2084      0.017     11.931      0.000
0.174       0.243
LotArea                       0.1153      0.014      8.314      0.000
0.088       0.143
GarageArea                    0.1095      0.013      8.463      0.000
```

```
0.084       0.135
BsmtQual_Ex                 0.2684      0.054      4.930      0.000
0.162       0.375
TotRmsAbvGrd                0.1492      0.014     10.523      0.000
0.121       0.177
BsmtFinType1_Unf           -0.0986      0.027     -3.708      0.000
-0.151      -0.046
YearRemodAdd                0.1527      0.020      7.618      0.000
0.113       0.192
Fireplaces                  0.0697      0.011      6.155      0.000
0.047       0.092
Neighborhood_NoRidge        0.4827      0.063      7.697      0.000
0.360       0.606
KitchenQual_Ex              0.2443      0.054      4.489      0.000
0.138       0.351
BsmtExposure_Gd             0.2184      0.038      5.794      0.000
0.144       0.292
Neighborhood_Crawfor        0.3447      0.055      6.210      0.000
0.236       0.454
KitchenAbvGr               -0.0528      0.012     -4.369      0.000
-0.077      -0.029
Neighborhood_StoneBr        0.4895      0.077      6.397      0.000
0.339       0.640
FullBath                    0.0928      0.015      6.032      0.000
0.063       0.123
Exterior1st_BrkFace         0.3889      0.055      7.066      0.000
0.281       0.497
Neighborhood_NridgHt        0.1454      0.053      2.759      0.006
0.042       0.249
SaleType_WD                -0.1431      0.029     -4.992      0.000
-0.199      -0.087
BsmtFullBath                0.0696      0.011      6.084      0.000
0.047       0.092
OpenPorchSF                 0.0386      0.011      3.520      0.000
0.017       0.060
MasVnrArea                  0.0484      0.014      3.581      0.000
0.022       0.075
Functional_Typ              0.1341      0.049      2.753      0.006
0.039       0.230
WoodDeckSF                  0.0438      0.011      4.141      0.000
0.023       0.065
SaleCondition_Abnorml      -0.1310      0.038     -3.475      0.001
-0.205      -0.057
KitchenQual_Gd              0.0344      0.027      1.276      0.202
-0.019       0.087
BldgType_Twnhs             -0.1459      0.061     -2.384      0.017
-0.266      -0.026
Foundation_BrkTil          -0.1597      0.038     -4.201      0.000
```

```
-0.234      -0.085
Neighborhood_BrkSide          0.2121      0.054      3.963      0.000
0.107       0.317
MSZoning_FV                   0.2071      0.060      3.441      0.001
0.089       0.325
OverallCond                   0.0437      0.013      3.459      0.001
0.019       0.069
Condition1_Norm               0.1023      0.030      3.450      0.001
0.044       0.160
MSZoning_Call                -0.5274      0.123     -4.279      0.000
-0.769      -0.286
GarageType_2Types            -0.2064      0.157     -1.311      0.190
-0.515       0.102
ScreenPorch                   0.0313      0.013      2.439      0.015
0.006       0.056
LotConfig_CulDSac             0.1124      0.039      2.917      0.004
0.037       0.188
GarageType_BuiltIn            0.0676      0.042      1.594      0.111
-0.016       0.151
HouseStyle_SLvl              -0.0988      0.045     -2.219      0.027
-0.186      -0.011
BsmtFinType1_GLQ              0.0439      0.030      1.456      0.146
-0.015       0.103
Exterior2nd_WdShng           -0.1314      0.061     -2.144      0.032
-0.252      -0.011
ExterQual_Ex                  0.0393      0.070      0.562      0.574
-0.098       0.177
LotConfig_FR2                -0.1410      0.050     -2.820      0.005
-0.239      -0.043
BsmtExposure_No              -0.0455      0.023     -1.966      0.049
-0.091      -0.000
HouseStyle_SFoyer            -0.1664      0.066     -2.529      0.012
-0.295      -0.037
SaleCondition_Alloca          0.2505      0.136      1.838      0.066
-0.017       0.518
MasVnrType_BrkCmn            -0.1903      0.090     -2.120      0.034
-0.366      -0.014
LotConfig_FR3                -0.2210      0.157     -1.411      0.159
-0.528       0.086
Neighborhood_NWAmes          -0.0669      0.046     -1.466      0.143
-0.156       0.023
HalfBath                      0.0351      0.011      3.094      0.002
0.013       0.057
Condition1_RRAe              -0.2186      0.105     -2.088      0.037
-0.424      -0.013
LandContour_Low              -0.0028      0.077     -0.036      0.971
-0.153       0.148
Exterior1st_WdSdng           -0.0603      0.060     -1.005      0.315
```

					-0.178	0.057
Functional_Min2	0.1626	0.078	2.097	0.036	0.010	0.315
Neighborhood_Gilbert	-0.0751	0.046	-1.627	0.104	-0.166	0.015
BsmtQual_Gd	0.0277	0.031	0.896	0.371	-0.033	0.088
Neighborhood_MeadowV	-0.3555	0.119	-2.990	0.003	-0.589	-0.122
Exterior2nd_CmentBd	0.5156	0.231	2.229	0.026	0.062	0.969
Neighborhood_Mitchel	-0.0981	0.055	-1.786	0.074	-0.206	0.010
LandSlope_Sev	-0.1418	0.151	-0.939	0.348	-0.438	0.155
Exterior2nd_WdSdng	0.0483	0.060	0.808	0.419	-0.069	0.166
Neighborhood_NAmes	-0.0444	0.033	-1.347	0.178	-0.109	0.020
Neighborhood_Edwards	-0.1134	0.043	-2.623	0.009	-0.198	-0.029
MSZoning_RL	0.1149	0.036	3.202	0.001	0.044	0.185
Condition1_PosN	0.0561	0.084	0.667	0.505	-0.109	0.221
MasVnrType_Stone	0.0234	0.036	0.649	0.517	-0.047	0.094
Exterior1st_Stone	0.4468	0.228	1.956	0.051	-0.001	0.895
Neighborhood_Timber	-0.0822	0.060	-1.374	0.170	-0.200	0.035
Exterior1st_CemntBd	-0.3979	0.228	-1.744	0.081	-0.846	0.050
GarageFinish_RFn	-0.0046	0.021	-0.218	0.827	-0.046	0.037
MSZoning_RH	0.1289	0.101	1.275	0.203	-0.069	0.327
EnclosedPorch	0.0170	0.010	1.750	0.080	-0.002	0.036
OverallQual:GarageArea	0.0240	0.018	1.348	0.178	-0.011	0.059
OverallQual:LotArea	0.0574	0.017	3.383	0.001	0.024	0.091
OverallQual:TotRmsAbvGrd	-0.0055	0.017	-0.329	0.742	-0.038	0.027
OverallQual:YearRemodAdd	0.0405	0.017	2.321	0.020	0.006	0.075
OverallQual:Fireplaces	0.0219	0.015	1.436	0.151		

```
-0.008       0.052
OverallQual:FullBath          0.0004       0.018       0.024       0.981
-0.035       0.036
OverallQual:BsmtFullBath      0.0185       0.015       1.255       0.210
-0.010       0.047
OverallQual:OpenPorchSF       0.0079       0.014       0.550       0.583
-0.020       0.036
OverallQual:MasVnrArea        0.0136       0.017       0.806       0.420
-0.020       0.047
OverallQual:WoodDeckSF        0.0004       0.015       0.026       0.979
-0.029       0.030
OverallQual:OverallCond       0.0076       0.014       0.534       0.593
-0.020       0.036
OverallQual:ScreenPorch       0.0191       0.014       1.335       0.182
-0.009       0.047
GarageArea:LotArea            0.0064       0.013       0.502       0.615
-0.019       0.031
GarageArea:TotRmsAbvGrd      -0.0102       0.016      -0.660       0.509
-0.041       0.020
GarageArea:YearRemodAdd       0.0193       0.015       1.261       0.208
-0.011       0.049
GarageArea:Fireplaces        -0.0059       0.015      -0.384       0.701
-0.036       0.024
GarageArea:FullBath           0.0161       0.018       0.897       0.370
-0.019       0.051
GarageArea:BsmtFullBath       0.0161       0.012       1.312       0.190
-0.008       0.040
GarageArea:OpenPorchSF       -0.0064       0.013      -0.477       0.634
-0.033       0.020
GarageArea:MasVnrArea        -0.0137       0.013      -1.050       0.294
-0.039       0.012
GarageArea:WoodDeckSF         0.0093       0.013       0.731       0.465
-0.016       0.034
GarageArea:OverallCond       -0.0028       0.012      -0.235       0.814
-0.026       0.020
GarageArea:ScreenPorch        0.0131       0.015       0.892       0.372
-0.016       0.042
LotArea:TotRmsAbvGrd          0.0043       0.015       0.299       0.765
-0.024       0.033
LotArea:YearRemodAdd         -0.0145       0.015      -0.971       0.332
-0.044       0.015
LotArea:Fireplaces           -0.0143       0.012      -1.207       0.227
-0.038       0.009
LotArea:FullBath             -0.0257       0.015      -1.675       0.094
-0.056       0.004
LotArea:BsmtFullBath         -0.0161       0.012      -1.291       0.197
-0.040       0.008
LotArea:OpenPorchSF           0.0168       0.012       1.432       0.152
```

					-0.006	0.040
LotArea:MasVnrArea	0.0049	0.012	0.419	0.675	-0.018	0.028
LotArea:WoodDeckSF	0.0004	0.009	0.049	0.961	-0.017	0.018
LotArea:OverallCond	-0.0020	0.012	-0.176	0.860	-0.025	0.021
LotArea:ScreenPorch	0.0045	0.011	0.423	0.672	-0.016	0.025
TotRmsAbvGrd:YearRemodAdd	0.0110	0.016	0.710	0.478	-0.019	0.042
TotRmsAbvGrd:Fireplaces	0.0192	0.014	1.372	0.170	-0.008	0.047
TotRmsAbvGrd:FullBath	0.0316	0.014	2.334	0.020	0.005	0.058
TotRmsAbvGrd:BsmtFullBath	0.0034	0.013	0.266	0.790	-0.022	0.029
TotRmsAbvGrd:OpenPorchSF	-0.0082	0.011	-0.762	0.446	-0.029	0.013
TotRmsAbvGrd:MasVnrArea	0.0124	0.013	0.932	0.352	-0.014	0.039
TotRmsAbvGrd:WoodDeckSF	-0.0214	0.012	-1.811	0.070	-0.045	0.002
TotRmsAbvGrd:OverallCond	0.0013	0.013	0.104	0.918	-0.024	0.027
TotRmsAbvGrd:ScreenPorch	-0.0153	0.011	-1.353	0.176	-0.037	0.007
YearRemodAdd:Fireplaces	-0.0110	0.013	-0.841	0.400	-0.037	0.015
YearRemodAdd:FullBath	0.0178	0.017	1.058	0.290	-0.015	0.051
YearRemodAdd:BsmtFullBath	0.0065	0.013	0.494	0.621	-0.019	0.032
YearRemodAdd:OpenPorchSF	0.0173	0.014	1.220	0.223	-0.010	0.045
YearRemodAdd:MasVnrArea	0.0112	0.018	0.630	0.529	-0.024	0.046
YearRemodAdd:WoodDeckSF	0.0120	0.014	0.863	0.388	-0.015	0.039
YearRemodAdd:OverallCond	-0.0377	0.011	-3.396	0.001	-0.060	-0.016
YearRemodAdd:ScreenPorch	-0.0021	0.012	-0.170	0.865	-0.027	0.022
Fireplaces:FullBath	0.0015	0.015	0.104	0.917	-0.027	0.030
Fireplaces:BsmtFullBath	-0.0019	0.011	-0.183	0.854	-0.023	0.019
Fireplaces:OpenPorchSF	-0.0007	0.012	-0.057	0.954		

					-0.024	0.023
Fireplaces:MasVnrArea	0.0124	0.012	0.996	0.319	-0.012	0.037
Fireplaces:WoodDeckSF	0.0251	0.011	2.302	0.021	0.004	0.047
Fireplaces:OverallCond	0.0329	0.011	2.968	0.003	0.011	0.055
Fireplaces:ScreenPorch	-0.0062	0.010	-0.597	0.551	-0.027	0.014
FullBath:BsmtFullBath	0.0015	0.013	0.119	0.905	-0.024	0.027
FullBath:OpenPorchSF	0.0048	0.014	0.354	0.723	-0.022	0.032
FullBath:MasVnrArea	-0.0073	0.015	-0.478	0.633	-0.037	0.023
FullBath:WoodDeckSF	0.0004	0.014	0.031	0.975	-0.027	0.028
FullBath:OverallCond	0.0008	0.013	0.058	0.954	-0.025	0.027
FullBath:ScreenPorch	0.0010	0.012	0.085	0.933	-0.022	0.024
BsmtFullBath:OpenPorchSF	-0.0042	0.011	-0.391	0.696	-0.025	0.017
BsmtFullBath:MasVnrArea	0.0082	0.012	0.704	0.482	-0.015	0.031
BsmtFullBath:WoodDeckSF	0.0067	0.010	0.666	0.505	-0.013	0.026
BsmtFullBath:OverallCond	0.0074	0.011	0.681	0.496	-0.014	0.029
BsmtFullBath:ScreenPorch	-0.0064	0.011	-0.588	0.557	-0.028	0.015
OpenPorchSF:MasVnrArea	-0.0041	0.011	-0.363	0.717	-0.026	0.018
OpenPorchSF:WoodDeckSF	0.0199	0.011	1.860	0.063	-0.001	0.041
OpenPorchSF:OverallCond	-0.0053	0.009	-0.564	0.573	-0.024	0.013
OpenPorchSF:ScreenPorch	0.0081	0.007	1.241	0.215	-0.005	0.021
MasVnrArea:WoodDeckSF	-0.0143	0.011	-1.354	0.176	-0.035	0.006
MasVnrArea:OverallCond	-0.0262	0.014	-1.926	0.054	-0.053	0.000
MasVnrArea:ScreenPorch	-0.0166	0.009	-1.929	0.054	-0.033	0.000
WoodDeckSF:OverallCond	-0.0014	0.010	-0.139	0.889	-0.021	0.018
WoodDeckSF:ScreenPorch	-0.0100	0.014	-0.736	0.462		

```
-0.037       0.017
OverallCond:ScreenPorch       0.0123      0.010      1.189      0.235
-0.008       0.033
==============================================================================
Omnibus:                      162.820   Durbin-Watson:                   1.922
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1130.673
Skew:                           0.332   Prob(JB):                    3.00e-246
Kurtosis:                       7.494   Cond. No.                         153.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

7.5 optimal model with interaction terms

[388]:
```python
print(optimal_linear_model_interaction.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              SalePrice   R-squared:                       0.919
Model:                            OLS   Adj. R-squared:                  0.912
Method:                 Least Squares   F-statistic:                     137.9
Date:                Thu, 07 May 2020   Prob (F-statistic):               0.00
Time:                        21:26:33   Log-Likelihood:                -213.67
No. Observations:                1315   AIC:                             629.3
Df Residuals:                    1214   BIC:                             1153.
Df Model:                         100
Covariance Type:            nonrobust
================================================================================
============
                              coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
------------
Intercept                  -0.3356      0.069     -4.834      0.000      -0.472
-0.199
OverallQual                 0.2092      0.016     12.809      0.000       0.177
0.241
LotArea                     0.1149      0.013      9.121      0.000       0.090
0.140
GarageArea                  0.1078      0.012      9.065      0.000       0.084
0.131
BsmtQual_Ex                 0.2696      0.052      5.222      0.000       0.168
0.371
TotRmsAbvGrd                0.1473      0.013     11.046      0.000       0.121
0.173
BsmtFinType1_Unf           -0.0976      0.026     -3.812      0.000      -0.148
```

						-0.047
YearRemodAdd	0.1497	0.018	8.128	0.000	0.114	0.186
Fireplaces	0.0744	0.010	7.136	0.000	0.054	0.095
Neighborhood_NoRidge	0.4834	0.059	8.182	0.000	0.368	0.599
KitchenQual_Ex	0.2427	0.052	4.705	0.000	0.142	0.344
BsmtExposure_Gd	0.2147	0.036	5.928	0.000	0.144	0.286
Neighborhood_Crawfor	0.3360	0.052	6.476	0.000	0.234	0.438
KitchenAbvGr	-0.0538	0.011	-4.972	0.000	-0.075	-0.033
Neighborhood_StoneBr	0.4948	0.071	6.983	0.000	0.356	0.634
FullBath	0.0921	0.014	6.441	0.000	0.064	0.120
Exterior1st_BrkFace	0.3837	0.052	7.315	0.000	0.281	0.487
Neighborhood_NridgHt	0.1506	0.050	2.987	0.003	0.052	0.249
SaleType_WD	-0.1425	0.028	-5.162	0.000	-0.197	-0.088
BsmtFullBath	0.0691	0.011	6.245	0.000	0.047	0.091
OpenPorchSF	0.0376	0.010	3.720	0.000	0.018	0.057
MasVnrArea	0.0475	0.012	3.994	0.000	0.024	0.071
Functional_Typ	0.1335	0.047	2.872	0.004	0.042	0.225
WoodDeckSF	0.0449	0.010	4.677	0.000	0.026	0.064
SaleCondition_Abnorml	-0.1266	0.037	-3.462	0.001	-0.198	-0.055
KitchenQual_Gd	0.0368	0.026	1.418	0.156	-0.014	0.088
BldgType_Twnhs	-0.1445	0.058	-2.498	0.013	-0.258	-0.031
Foundation_BrkTil	-0.1583	0.036	-4.397	0.000	-0.229	-0.088
Neighborhood_BrkSide	0.2187	0.051	4.286	0.000	0.119	0.319
MSZoning_FV	0.2086	0.056	3.746	0.000	0.099	0.318
OverallCond	0.0419	0.011	3.748	0.000	0.020	

						0.064
Condition1_Norm	0.1021	0.028	3.589	0.000	0.046	0.158
MSZoning_Call	-0.5239	0.118	-4.441	0.000	-0.755	-0.292
GarageType_2Types	-0.1375	0.140	-0.980	0.327	-0.413	0.138
ScreenPorch	0.0315	0.010	3.239	0.001	0.012	0.051
LotConfig_CulDSac	0.1158	0.037	3.127	0.002	0.043	0.188
GarageType_BuiltIn	0.0679	0.040	1.684	0.092	-0.011	0.147
HouseStyle_SLvl	-0.1076	0.042	-2.543	0.011	-0.191	-0.025
BsmtFinType1_GLQ	0.0385	0.028	1.354	0.176	-0.017	0.094
Exterior2nd_WdShng	-0.1358	0.059	-2.284	0.023	-0.252	-0.019
ExterQual_Ex	0.0597	0.064	0.937	0.349	-0.065	0.185
LotConfig_FR2	-0.1336	0.048	-2.771	0.006	-0.228	-0.039
BsmtExposure_No	-0.0434	0.022	-1.936	0.053	-0.087	0.001
HouseStyle_SFoyer	-0.1633	0.063	-2.577	0.010	-0.288	-0.039
SaleCondition_Alloca	0.2447	0.124	1.970	0.049	0.001	0.488
MasVnrType_BrkCmn	-0.2121	0.086	-2.479	0.013	-0.380	-0.044
LotConfig_FR3	-0.2301	0.153	-1.502	0.133	-0.531	0.070
Neighborhood_NWAmes	-0.0755	0.043	-1.743	0.082	-0.161	0.010
HalfBath	0.0335	0.011	3.089	0.002	0.012	0.055
Condition1_RRAe	-0.2150	0.101	-2.128	0.033	-0.413	-0.017
LandContour_Low	0.0135	0.072	0.187	0.852	-0.129	0.156
Exterior1st_WdSdng	-0.0757	0.057	-1.323	0.186	-0.188	0.037
Functional_Min2	0.1661	0.074	2.239	0.025	0.021	0.312
Neighborhood_Gilbert	-0.0854	0.044	-1.961	0.050	-0.171	6.07e-05
BsmtQual_Gd	0.0285	0.029	0.968	0.333	-0.029	

						0.086
Neighborhood_MeadowV	-0.3484	0.112	-3.115	0.002	-0.568	-0.129
Exterior2nd_CmentBd	0.4805	0.224	2.149	0.032	0.042	0.919
Neighborhood_Mitchel	-0.0922	0.053	-1.747	0.081	-0.196	0.011
LandSlope_Sev	-0.1837	0.137	-1.338	0.181	-0.453	0.086
Exterior2nd_WdSdng	0.0618	0.057	1.079	0.281	-0.051	0.174
Neighborhood_NAmes	-0.0469	0.032	-1.484	0.138	-0.109	0.015
Neighborhood_Edwards	-0.1137	0.042	-2.715	0.007	-0.196	-0.032
MSZoning_RL	0.1128	0.033	3.415	0.001	0.048	0.178
Condition1_PosN	0.0615	0.080	0.766	0.444	-0.096	0.219
MasVnrType_Stone	0.0214	0.035	0.619	0.536	-0.046	0.089
Exterior1st_Stone	0.4073	0.222	1.838	0.066	-0.027	0.842
Neighborhood_Timber	-0.0853	0.058	-1.480	0.139	-0.198	0.028
Exterior1st_CemntBd	-0.3576	0.221	-1.616	0.106	-0.792	0.077
GarageFinish_RFn	-0.0019	0.020	-0.093	0.926	-0.042	0.038
MSZoning_RH	0.1219	0.097	1.254	0.210	-0.069	0.312
EnclosedPorch	0.0153	0.009	1.639	0.101	-0.003	0.034
OverallQual:GarageArea	0.0201	0.012	1.678	0.094	-0.003	0.044
OverallQual:LotArea	0.0651	0.014	4.589	0.000	0.037	0.093
OverallQual:YearRemodAdd	0.0450	0.014	3.188	0.001	0.017	0.073
TotRmsAbvGrd:Fireplaces	0.0223	0.011	2.014	0.044	0.001	0.044
YearRemodAdd:OverallCond	-0.0352	0.010	-3.670	0.000	-0.054	-0.016
Fireplaces:OverallCond	0.0344	0.009	3.894	0.000	0.017	0.052
OverallQual:BsmtFullBath	0.0245	0.011	2.281	0.023	0.003	0.046
OpenPorchSF:WoodDeckSF	0.0193	0.010	2.017	0.044	0.001	

```
0.038
MasVnrArea:OverallCond      -0.0307      0.012     -2.598      0.009      -0.054
-0.008
TotRmsAbvGrd:FullBath        0.0315      0.010      3.091      0.002       0.012
0.052
LotArea:FullBath            -0.0220      0.010     -2.104      0.036      -0.043
-0.001
OverallQual:Fireplaces       0.0135      0.012      1.146      0.252      -0.010
0.037
MasVnrArea:ScreenPorch      -0.0136      0.008     -1.778      0.076      -0.029
0.001
YearRemodAdd:FullBath        0.0255      0.013      1.988      0.047       0.000
0.051
OverallQual:ScreenPorch      0.0202      0.011      1.911      0.056      -0.001
0.041
Fireplaces:WoodDeckSF        0.0224      0.010      2.359      0.018       0.004
0.041
TotRmsAbvGrd:WoodDeckSF     -0.0189      0.009     -2.037      0.042      -0.037
-0.001
LotArea:OpenPorchSF          0.0154      0.009      1.810      0.071      -0.001
0.032
LotArea:Fireplaces          -0.0146      0.010     -1.437      0.151      -0.035
0.005
Fireplaces:MasVnrArea        0.0188      0.010      1.799      0.072      -0.002
0.039
GarageArea:BsmtFullBath      0.0177      0.011      1.620      0.106      -0.004
0.039
LotArea:BsmtFullBath        -0.0129      0.010     -1.271      0.204      -0.033
0.007
GarageArea:YearRemodAdd      0.0220      0.013      1.695      0.090      -0.003
0.048
LotArea:YearRemodAdd        -0.0191      0.012     -1.534      0.125      -0.044
0.005
BsmtFullBath:ScreenPorch    -0.0095      0.009     -1.025      0.306      -0.028
0.009
YearRemodAdd:WoodDeckSF      0.0122      0.011      1.133      0.257      -0.009
0.033
YearRemodAdd:OpenPorchSF     0.0196      0.011      1.751      0.080      -0.002
0.042
OpenPorchSF:ScreenPorch      0.0092      0.005      1.737      0.083      -0.001
0.019
TotRmsAbvGrd:ScreenPorch    -0.0147      0.009     -1.683      0.093      -0.032
0.002
OverallCond:ScreenPorch      0.0102      0.009      1.107      0.268      -0.008
0.028
==============================================================================
Omnibus:                      158.514   Durbin-Watson:                   1.929
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1111.202
```

```
Skew:                           0.305   Prob(JB):                    5.08e-242
Kurtosis:                       7.462   Cond. No.                         111.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```