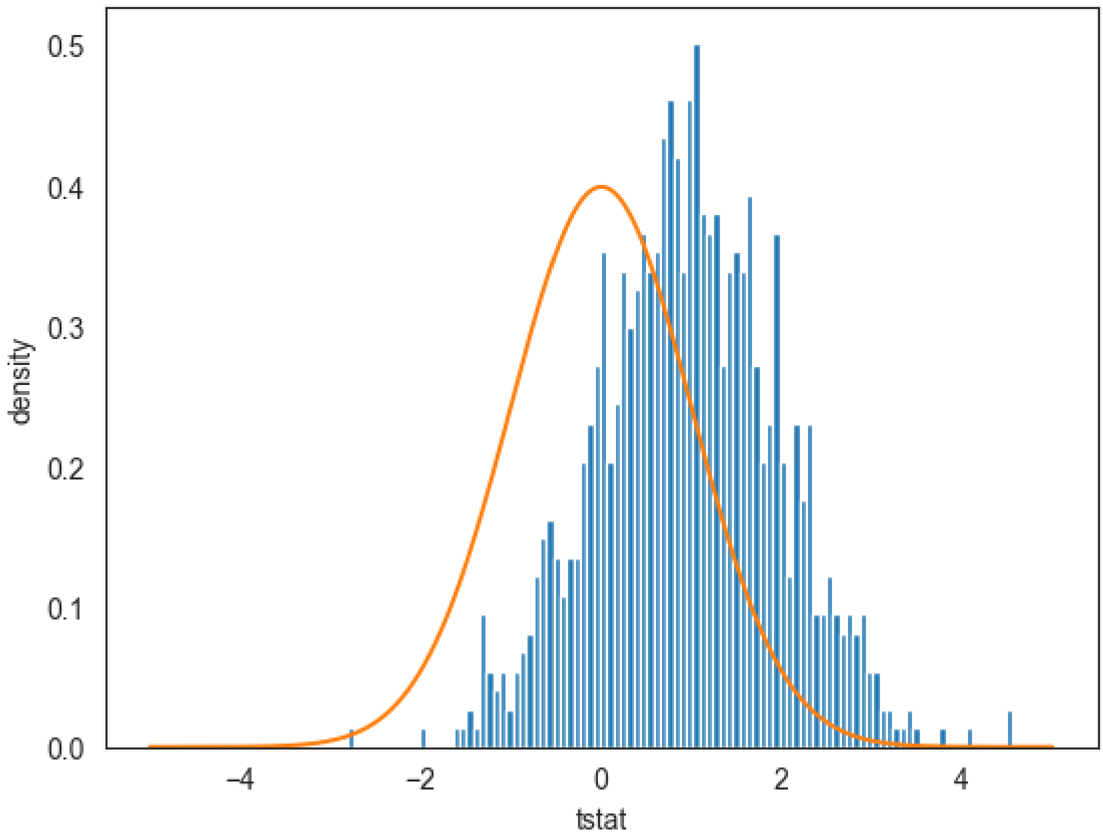
Econ-UB Assignment 2

Shalem Sumanthiran - sps9893

Theory: Simulating OVB

Question 1

In [ ]:
```
import numpy as np
import scipy.stats as sp
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import statsmodels.formula.api as smf
import statsmodels.api as sm
sns.set_style('white')

N=500
B=1000
beta1=[]
tstat=[]

np.random.seed(1)
for b in range(B):                                        #range iterates from 0 to 99
    x1=np.random.normal(0,1,N)                            # 1) generate artificial X1
    x2=0.1*x1+np.random.normal(0,1,N)                     # 1) generate artificial X2
    y=1.1*x1+0.5*x2+np.random.normal(0,1,N)               # 2) generate artificial Y
    rho=np.corrcoef(x1,y)                                 # store the correlation
    model = sm.OLS(y, x1)                                 # 3) regress Y on X1
    result = model.fit()

    beta1.append(result.params[0])
    tstat.append((beta1[b]-1.1)/(result.bse[0]))          # store the beta_hat


myhist=plt.hist(tstat,bins=100, density=True)
x_axis = np.arange(-5, 5, 0.001)
mynorm=plt.plot(x_axis, sp.norm.pdf(x_axis,0,1)) # Mean = 0, SD = 1
plt.xlabel('tstat')
plt.ylabel('density')
plt.show()
```



Omitted variable bias occurs under two conditions:

1. Z is a determinant of Y (i.e. Z is part of u)

2. Z is correlated with the regressor X

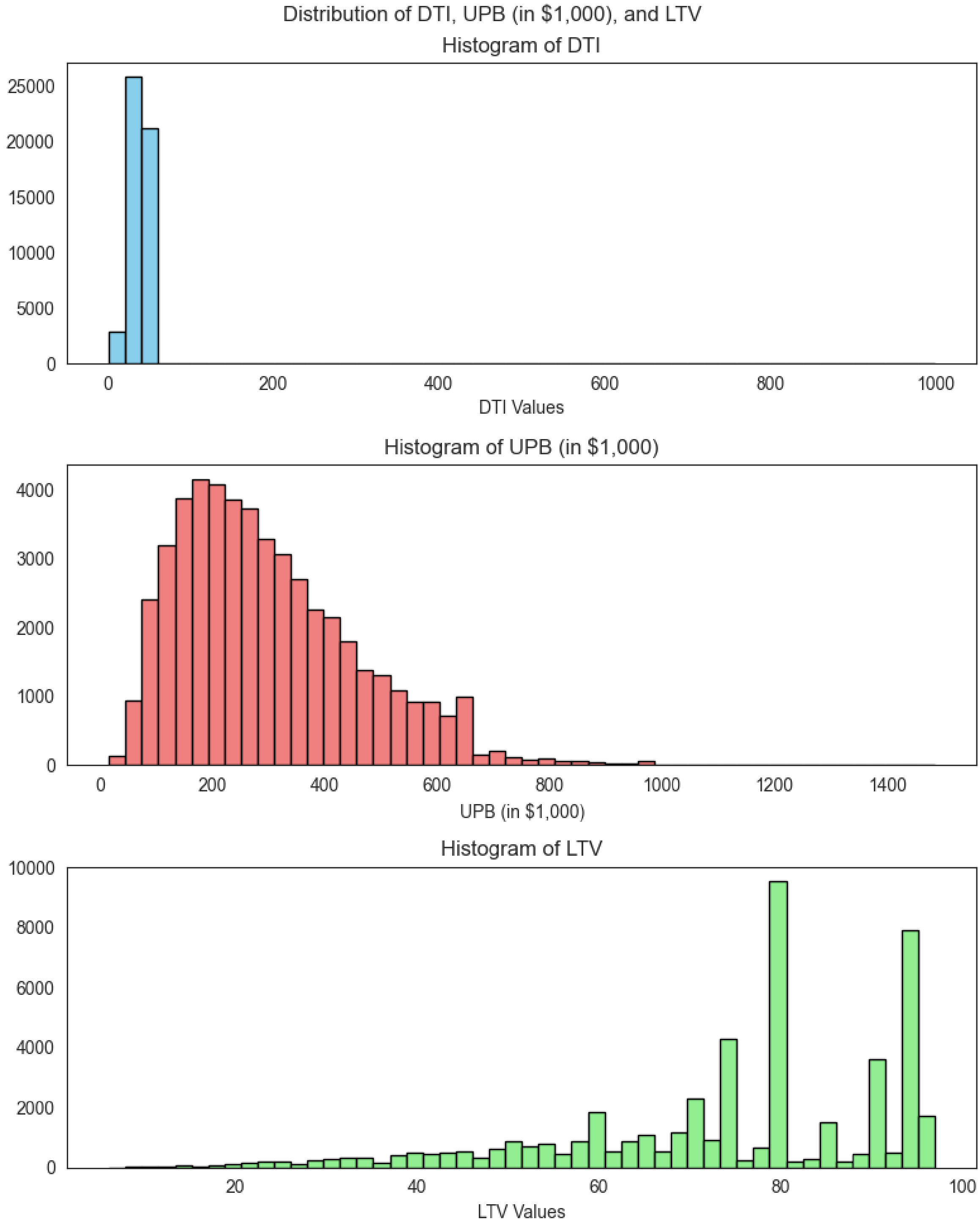
In this model, we observe Omitted Variable bias. The simulated X_2 is defined using X_1, which implies that Y_i is determined by the error term. Therefore, we can conclude that there is Omitted Variable bias as both conditions are satisfied. This is also seen in the graph, as the histogram does not align with the normal distribution.

Empirical

Question 2

In [ ]:
```
import pandas as pd
from scipy.stats import kurtosis, skew

var_names = ["Credit_Score","First_Payment_date","First_Time_Homebuyer", "Maturi
"MSA","Mortgage_Insurance_Percentage","Number_Units","Occupancy_Status","CLTV",
"DTI","UPB","LTV","Interest_Rate","Channel","Prepayment_Penalty",
"Amortization_Type","State","Property_Type","Postal_Code","Sequence_Number",
"Purpose","Loan_Term","Number_Borrowers","Seller_Name","Servicer_Name",
"Super_Conforming","Pre-HARP_Loan","Program_Indicator","HARP_Indicator",
"Valuation_Method","Interest_Only", "Insurance_cancellation"]

mysample = pd.read_table("C:/Users/shale/Downloads/sample_2022/sample_orig_2022.

for column in var_names:
    mysample = mysample[mysample[column] != 9999]
```

```
mysample['UPB'] = mysample['UPB'].div(1000)

figure, (ax0, ax1, ax2) = plt.subplots(nrows=3, ncols=1, figsize=(8, 10))

ax0.hist(mysample['DTI'], bins=50, color='skyblue', edgecolor='black')
ax0.set_title('Histogram of DTI')
ax0.set_xlabel('DTI Values')

ax1.hist(mysample['UPB'], bins=50, color='lightcoral', edgecolor='black')
ax1.set_title('Histogram of UPB (in $1,000)')
ax1.set_xlabel('UPB (in $1,000)')

ax2.hist(mysample['LTV'], bins=50, color='lightgreen', edgecolor='black')
ax2.set_title('Histogram of LTV')
ax2.set_xlabel('LTV Values')

plt.suptitle('Distribution of DTI, UPB (in $1,000), and LTV')
plt.tight_layout()
plt.show()

distribution_characteristics = mysample[['DTI', 'UPB', 'LTV']].describe()
distribution_characteristics.loc['kurtosis'] = [kurtosis(mysample['DTI']), kurto
distribution_characteristics.loc['skewness'] = [skew(mysample['DTI']), skew(mysa

print(distribution_characteristics)
```




```
                 DTI          UPB          LTV
count   49983.000000 49983.000000 49983.000000
mean       37.097773   299.661985    74.070624
std        15.225280   159.957928    18.483965
min         1.000000    16.000000     6.000000
25%        31.000000   176.000000    64.000000
50%        39.000000   270.000000    80.000000
75%        44.000000   394.000000    90.000000
max       999.000000  1485.000000    97.000000
kurtosis 2547.419542     0.741601     0.360242
skewness   40.205618     0.895555    -0.920109
```

The distribution characteristics above along with the histograms indicate the following:

For DTI - The distribution is right-skewed, indicating more observations relatively less than the mean, with one outlier of 999 given that the mean is 37.09 For UPB - The distribution is right-skewed, indicating more observations relatively less than the mean. For LTV - The distribution is left-skewed, indicating more observations relativey greater than the mean.

All other characteristics are displayed in the table above.

Based on economic reasoning, the variables should be pricing factors that determine the interst rate of the loan, based on the following:

Debt-to-Income Ratio (DTI): DTI is a measure of the borrower's ability to manage debt payments in relation to their income. A high DTI indicates that a borrower has a significant portion of their income allocated to debt payments. Borrowers with lower DTIs may get lower interest rates because they are considered lower risk.

Unpaid Principal Balance (UPB): The UPB represents the outstanding balance of the loan, which decreases over time as the borrower makes payments. Therefore it may indicate a lower ability to repay the loan on the part of the borrower if they have significant outstanding balance, signalling a higher risk.

Loan-to-Value Ratio (LTV): LTV measures the ratio of the loan amount to the appraised value of the collateral. Higher LTVs imply that the borrower is financing a larger portion of the purchase price, which can be considered riskier.

Question 3

In [ ]:
```
correlation_matrix = mysample[['Credit_Score', 'DTI', 'UPB', 'LTV']].corr()
print(correlation_matrix)
```

```
              Credit_Score       DTI       UPB       LTV
Credit_Score      1.000000 -0.052572  0.149520  0.062062
DTI              -0.052572  1.000000  0.070790  0.060250
UPB               0.149520  0.070790  1.000000  0.259811
LTV               0.062062  0.060250  0.259811  1.000000
```

There is unlikely to be a great risk of multicollinearity in the data, as we can see above that the respective correlations between each variable are relatively small (below 0.26 for any of them), therefore it should not be difficult to determine the individual effect of each independent variable on the dependent variable.

Question 4

In [ ]:
```
import statsmodels.formula.api as smf
results = smf.ols('Interest_Rate ~ Credit_Score + DTI + UPB + LTV', data=mysampl
print(results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Interest_Rate   R-squared:                       0.047
Model:                            OLS   Adj. R-squared:                  0.047
Method:                 Least Squares   F-statistic:                     531.0
Date:                Thu, 19 Oct 2023   Prob (F-statistic):               0.00
Time:                        13:55:01   Log-Likelihood:                -79920.
No. Observations:               49983   AIC:                         1.599e+05
Df Residuals:                   49978   BIC:                         1.599e+05
Df Model:                           4
Covariance Type:                  HC3
================================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept       5.0205      0.105     48.002      0.000       4.816       5.226
Credit_Score   -0.0013      0.000    -10.637      0.000      -0.002      -0.001
DTI             0.0044      0.001      3.878      0.000       0.002       0.007
UPB            -0.0005   3.59e-05    -15.229      0.000      -0.001      -0.000
LTV             0.0138      0.000     43.273      0.000       0.013       0.014
==============================================================================
Omnibus:                     2140.979   Durbin-Watson:                   0.397
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              912.298
Skew:                          -0.049   Prob(JB):                    7.89e-199
Kurtosis:                       2.345   Cond. No.                     1.35e+04
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)
[2] The condition number is large, 1.35e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The coefficient of the credit score indicates that when the Credit Score changes by one unit, the Interest Rate reduces by 0.0013 units on average ceteris paribus.

The coefficients of DTI, and LTV have the expected signs of being positive as per the earlier discussion of the expected impact of the variables, however the UPB variable has a negative sign which was not expected, which may indicate something wrong with the model or a need to analyze more carefully our expectations.

The impact of all variables are statistically significant at a 5% level, which indicates that the variables will have an impact on Interest Rate more than 95 times out of 100 if we were to repeat the regression with different samples.

In [ ]:

```
results2 = smf.ols('Interest_Rate ~ Credit_Score', data=mysample).fit(cov_type='
print(results2.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:          Interest_Rate   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                  0.002
Method:                 Least Squares   F-statistic:                     119.5
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           8.72e-28
Time:                        13:55:01   Log-Likelihood:                -81058.
No. Observations:               49983   AIC:                         1.621e+05
Df Residuals:                   49981   BIC:                         1.621e+05
Df Model:                           1                                         
Covariance Type:                  HC3                                         
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept        6.0532      0.089     68.023      0.000       5.879       6.228
Credit_Score    -0.0013      0.000    -10.932      0.000      -0.002      -0.001
==============================================================================
Omnibus:                     2954.745   Durbin-Watson:                   0.359
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1173.635
Skew:                          -0.113   Prob(JB):                    1.41e-255
Kurtosis:                       2.284   Cond. No.                     1.20e+04
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)
[2] The condition number is large, 1.2e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

As seen above, the R squared of the model with only Credit Score is 0.002 while the R squared of the model with all the variables is 0.047, indicating that the model with the higher R-squared value explains more of the variation of the data, making it more accurate.

The coefficient estimate of Credit Score did not change significantly when looking at the standard error of the effect of the variable on Interest Rate. In addition, the effect is similar as well.

Question 5

In [ ]:
```python
# Define the null hypothesis
hypothesis = '(DTI = 0, UPB = 0, LTV = 0)'

# Perform the Wald test for joint hypothesis
wald_test = results.wald_test(hypothesis)

# Get the test statistic and p-value
test_statistic = wald_test.statistic
p_value = wald_test.pvalue

# Define the significance level
alpha = 0.05

# Check if the joint hypothesis is rejected at the 5% significance level
if p_value < alpha:
    print("Reject the joint hypothesis at the 5% significance level.")
else:
    print("Fail to reject the joint hypothesis at the 5% significance level.")
```

```
# Print the test statistic and p-value
print("Test Statistic:", test_statistic)
print("P-Value:", p_value)
```

```
Reject the joint hypothesis at the 5% significance level.
Test Statistic: [[2008.28904245]]
P-Value: 0.0
```

```
C:\Users\shale\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2
kfra8p0\LocalCache\local-packages\Python311\site-packages\statsmodels\base\model.
py:1906: FutureWarning: The behavior of wald_test will change after 0.14 to retur
ning scalar test statistic values. To get the future behavior now, set scalar to
True. To silence this message while retaining the legacy behavior, set scalar to
False.
  warnings.warn(
```

Therefore we reject the joint hypothesis at the 5% significance level, indicating that our restricted and unrestricted models are different to each other.

Question 6

In [ ]:
```
results3 = smf.ols('Interest_Rate ~ Credit_Score + DTI + UPB + LTV + State', dat
print(results3.summary())
```

OLS Regression Results

Dep. Variable:	Interest_Rate	R-squared:	0.049
Model:	OLS	Adj. R-squared:	0.048
Method:	Least Squares	F-statistic:	39.49
Date:	Thu, 19 Oct 2023	Prob (F-statistic):	0.00
Time:	13:55:01	Log-Likelihood:	-79856.
No. Observations:	49983	AIC:	1.598e+05
Df Residuals:	49925	BIC:	1.603e+05
Df Model:	57		
Covariance Type:	HC3		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	4.8252	0.164	29.471	0.000	4.504	5.146
State[T.AL]	0.2007	0.135	1.488	0.137	-0.064	0.465
State[T.AR]	0.1244	0.142	0.875	0.381	-0.154	0.403
State[T.AZ]	0.1758	0.130	1.356	0.175	-0.078	0.430
State[T.CA]	0.1425	0.128	1.114	0.265	-0.108	0.393
State[T.CO]	0.1957	0.131	1.497	0.134	-0.060	0.452
State[T.CT]	0.1899	0.136	1.393	0.164	-0.077	0.457
State[T.DC]	0.1308	0.191	0.687	0.492	-0.243	0.504
State[T.DE]	0.2244	0.146	1.535	0.125	-0.062	0.511
State[T.FL]	0.2096	0.128	1.644	0.100	-0.040	0.460
State[T.GA]	0.2155	0.129	1.670	0.095	-0.037	0.468
State[T.GU]	-0.6965	0.748	-0.931	0.352	-2.163	0.770
State[T.HI]	0.0178	0.172	0.104	0.918	-0.320	0.356
State[T.IA]	0.0682	0.139	0.491	0.623	-0.204	0.340
State[T.ID]	0.3286	0.140	2.344	0.019	0.054	0.603
State[T.IL]	0.2480	0.129	1.921	0.055	-0.005	0.501
State[T.IN]	0.3651	0.131	2.792	0.005	0.109	0.621
State[T.KS]	0.2276	0.141	1.611	0.107	-0.049	0.504
State[T.KY]	0.2693	0.135	1.989	0.047	0.004	0.535
State[T.LA]	0.2875	0.137	2.099	0.036	0.019	0.556
State[T.MA]	0.1470	0.134	1.101	0.271	-0.115	0.409
State[T.MD]	0.0912	0.132	0.692	0.489	-0.167	0.350
State[T.ME]	0.1913	0.150	1.274	0.203	-0.103	0.486
State[T.MI]	0.3349	0.130	2.581	0.010	0.081	0.589
State[T.MN]	0.1879	0.131	1.438	0.150	-0.068	0.444
State[T.MO]	0.2695	0.132	2.048	0.041	0.012	0.527
State[T.MS]	0.1062	0.155	0.683	0.495	-0.199	0.411
State[T.MT]	0.3506	0.155	2.258	0.024	0.046	0.655
State[T.NC]	0.1532	0.129	1.184	0.236	-0.100	0.407
State[T.ND]	-0.0592	0.178	-0.332	0.740	-0.409	0.290
State[T.NE]	0.1454	0.147	0.989	0.323	-0.143	0.434
State[T.NH]	0.4129	0.148	2.796	0.005	0.123	0.702
State[T.NJ]	0.1673	0.131	1.278	0.201	-0.089	0.424
State[T.NM]	0.2931	0.148	1.986	0.047	0.004	0.582
State[T.NV]	0.1331	0.135	0.983	0.326	-0.132	0.399
State[T.NY]	0.1586	0.130	1.224	0.221	-0.095	0.413
State[T.OH]	0.2606	0.129	2.016	0.044	0.007	0.514
State[T.OK]	0.2561	0.137	1.869	0.062	-0.013	0.525
State[T.OR]	0.2265	0.133	1.700	0.089	-0.035	0.488
State[T.PA]	0.2185	0.129	1.692	0.091	-0.035	0.472
State[T.PR]	-0.8364	1.063	-0.787	0.432	-2.920	1.248
State[T.RI]	0.1071	0.169	0.634	0.526	-0.224	0.438
State[T.SC]	0.1823	0.132	1.384	0.166	-0.076	0.440
State[T.SD]	0.1111	0.165	0.672	0.502	-0.213	0.435
State[T.TN]	0.2021	0.131	1.547	0.122	-0.054	0.458
State[T.TX]	0.1848	0.127	1.451	0.147	-0.065	0.435

```
State[T.UT]       0.1714      0.133      1.287      0.198      -0.090       0.432
State[T.VA]       0.1539      0.130      1.181      0.238      -0.102       0.409
State[T.VI]      -0.5079      0.946     -0.537      0.591      -2.362       1.346
State[T.VT]       0.0956      0.174      0.550      0.582      -0.245       0.436
State[T.WA]       0.1947      0.131      1.489      0.137      -0.062       0.451
State[T.WI]       0.2018      0.134      1.507      0.132      -0.061       0.464
State[T.WV]       0.1121      0.161      0.698      0.485      -0.203       0.427
State[T.WY]       0.1148      0.172      0.668      0.504      -0.222       0.452
Credit_Score     -0.0013      0.000    -10.690      0.000      -0.002      -0.001
DTI               0.0045      0.001      3.862      0.000       0.002       0.007
UPB              -0.0004   4.01e-05    -11.208      0.000      -0.001      -0.000
LTV               0.0134      0.000     39.714      0.000       0.013       0.014
==============================================================================
Omnibus:                     2045.484   Durbin-Watson:                   0.400
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              888.226
Skew:                          -0.051   Prob(JB):                    1.33e-193
Kurtosis:                       2.355   Cond. No.                     1.49e+05
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)
[2] The condition number is large, 1.49e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The state GA and IL have statistically significant positive coefficients at the 10% significance level, indicating that because the loans originate from those states, the Interest Rate will increase by 0.2155 and 0.2480 respectively on average ceteris paribus.

The inclusion of state did not change the impact of Credit Score on Interest Rate significantly, as the standard errors and coefficient are still the same.

Question 7

In [ ]:

```
mysample['Log_Credit_Score'] = np.log(mysample['Credit_Score'])
mysample['sqr_LTV'] = np.square(mysample['LTV'])

results4 = smf.ols('Interest_Rate ~ Log_Credit_Score + DTI + UPB + State + sqr_L
print(results4.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Interest_Rate   R-squared:                       0.055
Model:                            OLS   Adj. R-squared:                  0.053
Method:                 Least Squares   F-statistic:                     46.91
Date:                Thu, 19 Oct 2023   Prob (F-statistic):               0.00
Time:                        13:55:02   Log-Likelihood:                -79717.
No. Observations:               49983   AIC:                         1.596e+05
Df Residuals:                   49925   BIC:                         1.601e+05
Df Model:                          57
Covariance Type:                  HC3
================================================================================
===
                       coef    std err          z      P>|z|      [0.025      0.9
75]
--------------------------------------------------------------------------------
---
Intercept           11.0304      0.601     18.344      0.000       9.852      12.
209
State[T.AL]          0.2085      0.134      1.555      0.120      -0.054       0.
471
State[T.AR]          0.1362      0.141      0.964      0.335      -0.141       0.
413
State[T.AZ]          0.1850      0.129      1.437      0.151      -0.067       0.
437
State[T.CA]          0.1501      0.127      1.181      0.238      -0.099       0.
399
State[T.CO]          0.2043      0.130      1.574      0.115      -0.050       0.
459
State[T.CT]          0.1966      0.135      1.452      0.147      -0.069       0.
462
State[T.DC]          0.1246      0.189      0.660      0.509      -0.245       0.
495
State[T.DE]          0.2314      0.145      1.593      0.111      -0.053       0.
516
State[T.FL]          0.2198      0.127      1.736      0.083      -0.028       0.
468
State[T.GA]          0.2219      0.128      1.731      0.083      -0.029       0.
473
State[T.GU]         -0.6432      0.749     -0.859      0.390      -2.111       0.
824
State[T.HI]          0.0239      0.172      0.139      0.889      -0.313       0.
361
State[T.IA]          0.0692      0.138      0.502      0.616      -0.201       0.
339
State[T.ID]          0.3356      0.139      2.409      0.016       0.063       0.
609
State[T.IL]          0.2496      0.128      1.947      0.052      -0.002       0.
501
State[T.IN]          0.3665      0.130      2.823      0.005       0.112       0.
621
State[T.KS]          0.2295      0.140      1.635      0.102      -0.046       0.
505
State[T.KY]          0.2718      0.135      2.021      0.043       0.008       0.
535
State[T.LA]          0.2915      0.136      2.141      0.032       0.025       0.
558
State[T.MA]          0.1533      0.133      1.155      0.248      -0.107       0.
413
State[T.MD]          0.0928      0.131      0.709      0.478      -0.164       0.
```

```
349
State[T.ME]       0.1964      0.149      1.315      0.188     -0.096       0.
489
State[T.MI]       0.3351      0.129      2.600      0.009      0.083       0.
588
State[T.MN]       0.1869      0.130      1.441      0.150     -0.067       0.
441
State[T.MO]       0.2733      0.131      2.091      0.037      0.017       0.
529
State[T.MS]       0.1200      0.154      0.777      0.437     -0.183       0.
423
State[T.MT]       0.3494      0.155      2.259      0.024      0.046       0.
653
State[T.NC]       0.1605      0.129      1.249      0.212     -0.091       0.
412
State[T.ND]      -0.0653      0.177     -0.369      0.712     -0.412       0.
282
State[T.NE]       0.1468      0.146      1.004      0.315     -0.140       0.
433
State[T.NH]       0.4199      0.147      2.861      0.004      0.132       0.
708
State[T.NJ]       0.1779      0.130      1.368      0.171     -0.077       0.
433
State[T.NM]       0.3055      0.147      2.083      0.037      0.018       0.
593
State[T.NV]       0.1397      0.135      1.039      0.299     -0.124       0.
403
State[T.NY]       0.1656      0.129      1.287      0.198     -0.087       0.
418
State[T.OH]       0.2638      0.128      2.055      0.040      0.012       0.
515
State[T.OK]       0.2581      0.136      1.895      0.058     -0.009       0.
525
State[T.OR]       0.2300      0.132      1.738      0.082     -0.029       0.
489
State[T.PA]       0.2223      0.128      1.733      0.083     -0.029       0.
474
State[T.PR]      -0.8089      1.067     -0.758      0.448     -2.899       1.
282
State[T.RI]       0.1187      0.168      0.706      0.480     -0.211       0.
448
State[T.SC]       0.1901      0.131      1.454      0.146     -0.066       0.
446
State[T.SD]       0.1149      0.165      0.698      0.485     -0.208       0.
438
State[T.TN]       0.2117      0.130      1.632      0.103     -0.043       0.
466
State[T.TX]       0.1917      0.127      1.515      0.130     -0.056       0.
440
State[T.UT]       0.1778      0.132      1.344      0.179     -0.081       0.
437
State[T.VA]       0.1587      0.129      1.226      0.220     -0.095       0.
412
State[T.VI]      -0.5815      0.995     -0.585      0.559     -2.531       1.
368
State[T.VT]       0.1115      0.173      0.646      0.518     -0.227       0.
450
State[T.WA]       0.2000      0.130      1.540      0.124     -0.055       0.
455
State[T.WI]       0.2045      0.133      1.538      0.124     -0.056       0.
```

```
465
State[T.WV]          0.1130      0.160      0.707      0.480      -0.200       0.
426
State[T.WY]          0.1181      0.171      0.690      0.490      -0.218       0.
454
Log_Credit_Score    -1.0300      0.088    -11.707      0.000      -1.202      -0.
858
DTI                  0.0044      0.001      3.888      0.000       0.002       0.
007
UPB                 -0.0004   3.98e-05    -11.232      0.000      -0.001      -0.
000
sqr_LTV              0.0001   2.42e-06     44.179      0.000       0.000       0.
000
==============================================================================
Omnibus:                     2033.639   Durbin-Watson:                   0.410
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              881.034
Skew:                          -0.044   Prob(JB):                    4.85e-192
Kurtosis:                       2.356   Cond. No.                     1.17e+06
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)
[2] The condition number is large, 1.17e+06. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The effect of the log of credit score on interest rate indicates that for each percentage change in the Credit Score, the Interest Rate reduces by 1.0300/1000.

The effect of the square of the LTV on Interest Rate is difficult to interpret accurately as it is not linear and depends on the individual values of LTV.

We should use this model as opposed to the previous one as the adjusted R-squared indicates that this model explains the variability in the data better as 0.053 > 0.048.