In [1]:
```
#Violent Crime Category
```

In [2]:
```
import pandas as pd
import numpy as np
import statsmodels.api as sm
from statsmodels.stats.outliers_influence import variance_inflation_factor
from linearmodels.panel import PanelOLS
from linearmodels.panel import compare
```

In [3]:
```
df = pd.read_csv('panel_data/violent new.csv')
#df['l_ipc'] = np.log(df['ipc_cr'])
df.head()
```

Out[3]:

	s.no.	districts	year	type	violent_crimes	pop_in_lak	violent_cr	avg_temp	tot
0	1	ariyalur	2011	violent crimes	148	7.52	19.67	28.312353	1103.207
1	1	ariyalur	2012	violent crimes	176	7.63	23.05	28.777312	973.207
2	1	ariyalur	2013	violent crime	155	7.76	19.97	28.730311	870.158
3	1	ariyalur	2014	violent crime	127	7.88	16.11	28.536042	1090.802
4	1	ariyalur	2015	violent crime	85	8.00	10.60	28.565911	1501.644

In [4]:
```
df = df.set_index(['districts','year'])
y = df['violent_cr']
X = df[['avg_temp','tot_rf']]
```

In [5]:
```
#PooledOLS Estimation
X = sm.add_constant(X)
pols = PanelOLS(y,X)
pols_result = pols.fit()
print(pols_result.summary)
```

```
                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:             violent_cr   R-squared:                        0.0430
Estimator:                   PanelOLS   R-squared (Between):              0.0967
No. Observations:                 384   R-squared (Within):              -0.0184
Date:                Wed, Nov 12 2025   R-squared (Overall):              0.0430
Time:                        18:28:44   Log-likelihood                   -1289.0
Cov. Estimator:            Unadjusted
                                        F-statistic:                      8.5613
Entities:                          32   P-value                           0.0002
Avg Obs:                       12.000   Distribution:                   F(2,381)
Min Obs:                       12.000
Max Obs:                       12.000   F-statistic (robust):             8.5613
                                        P-value                           0.0002
Time periods:                      12   Distribution:                   F(2,381)
Avg Obs:                       32.000
Min Obs:                       32.000
Max Obs:                       32.000

                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
const          7.1199     3.5388     2.0119     0.0449      0.1619      14.078
avg_temp       0.3714     0.1125     3.3021     0.0011      0.1503      0.5926
tot_rf        -0.0008     0.0008    -0.9385     0.3486     -0.0024      0.0009
==============================================================================
```

In [6]:

```
#FE Model Estimation
X = sm.add_constant(X)
FEmodel = PanelOLS(y,X,entity_effects=True)
feresult = FEmodel.fit()
print(feresult.summary)
```

```
                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:             violent_cr   R-squared:                        0.0027
Estimator:                   PanelOLS   R-squared (Between):             -0.1497
No. Observations:                 384   R-squared (Within):               0.0027
Date:                Wed, Nov 12 2025   R-squared (Overall):             -0.0786
Time:                        18:28:44   Log-likelihood                   -1150.5
Cov. Estimator:            Unadjusted
                                        F-statistic:                      0.4738
Entities:                          32   P-value                           0.6230
Avg Obs:                       12.000   Distribution:                   F(2,350)
Min Obs:                       12.000
Max Obs:                       12.000   F-statistic (robust):             0.4738
                                        P-value                           0.6230
Time periods:                      12   Distribution:                   F(2,350)
Avg Obs:                       32.000
Min Obs:                       32.000
Max Obs:                       32.000

                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
const          23.220     12.266     1.8931     0.0592     -0.9031      47.344
avg_temp      -0.2709     0.4392    -0.6168     0.5378     -1.1348      0.5930
tot_rf         0.0003     0.0008     0.4033     0.6870     -0.0012      0.0018
==============================================================================

F-test for Poolability: 11.930
P-value: 0.0000
Distribution: F(31,350)

Included effects: Entity
```

In [7]:

```
#RE Model Estimation
from linearmodels.panel import RandomEffects
import statsmodels.api as sm
X = sm.add_constant(X)
REmodel = RandomEffects(y,X)
reresult = REmodel.fit()
print(reresult.summary)
```

```
                        RandomEffects Estimation Summary
================================================================================
Dep. Variable:             violent_cr   R-squared:                        0.0033
Estimator:              RandomEffects   R-squared (Between):              0.0549
No. Observations:                 384   R-squared (Within):              -0.0016
Date:                Wed, Nov 12 2025   R-squared (Overall):              0.0285
Time:                        18:28:44   Log-likelihood                   -1167.7
Cov. Estimator:            Unadjusted
                                        F-statistic:                      0.6212
Entities:                          32   P-value                           0.5379
Avg Obs:                       12.000   Distribution:                   F(2,381)
Min Obs:                       12.000
Max Obs:                       12.000   F-statistic (robust):             0.6212
                                        P-value                           0.5379
Time periods:                      12   Distribution:                   F(2,381)
Avg Obs:                       32.000
Min Obs:                       32.000
Max Obs:                       32.000

                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
const          8.9686     6.7143     1.3357     0.1824     -4.2332      22.170
avg_temp       0.2500     0.2346     1.0658     0.2872     -0.2112      0.7113
tot_rf         0.0005     0.0007     0.6711     0.5026     -0.0009      0.0019
==============================================================================
```

In [8]:

```python
#Hausman Test
from numpy.linalg import inv
from scipy.stats import chi2

b_FE = feresult.params
b_RE = reresult.params

common_coef = list(set(b_FE.index) & set(b_RE.index))

if 'const' in common_coef:
    common_coef.remove('const')

b_FE = b_FE[common_coef]
b_RE = b_RE[common_coef]

V_FE = feresult.cov
V_RE = reresult.cov

diff = b_FE - b_RE
diff_var = V_FE.loc[common_coef, common_coef] - V_RE.loc[common_coef, common_coef]

hausman_stat = np.dot(np.dot(diff.T, inv(diff_var)), diff)

df_h = len(diff)
p_value = 1 - chi2.cdf(hausman_stat, df_h)

print("Hausman Test Statistic:", round(hausman_stat, 3))
print("Degrees of Freedom:", df_h)
```

```python
print("p-value:", round(p_value, 4))
```

```
Hausman Test Statistic: 4.143
Degrees of Freedom: 2
p-value: 0.126
```

In [9]:
```python
#Diagnostic Checks
from statsmodels.stats.diagnostic import het_breuschpagan, het_white
from statsmodels.stats.stattools import durbin_watson
```

In [10]:
```python
#Test for Heteroskedasticity

#H0: No heteroskedasticity
#H1: Heteroskedasticity exists

#p-value <= 0.05 ---> Heteroskedasticity; p-value > 0.05 ---> Homoskedasticity

print('Breusch-Pagan Test')
residuals = reresult.resids
bp_test = het_breuschpagan(residuals, X)
bp_labels = ['Lagrange multiplier statistic','p-value','f-value','f p-value']
print(dict(zip(bp_labels, bp_test)))
print()
print('White Test')
white_test = het_white(residuals,X)
white_labels = ['LM stat','LM p-value','F p-value']
print(dict(zip(white_labels,white_test)))
```

```
Breusch-Pagan Test
{'Lagrange multiplier statistic': np.float64(4.210322820038144), 'p-value': np.float
64(0.12182600874362186), 'f-value': np.float64(2.1118701887129188), 'f p-value': n
p.float64(0.12242582224810573)}

White Test
{'LM stat': np.float64(13.773621666102258), 'LM p-value': np.float64(0.0171131813839
34493), 'F p-value': np.float64(2.8125651193287555)}
```

In [11]:
```python
#Test for serial correlation (autocorrelation)

#Durbin-Watson statistic ranges between 0 to 4

#DW statistic = 2 ---> No autocorrelation
#DW statistic < 2 ---> Positive autocorrelation
#DW statistic > 2 ---> Negative autocorrelation

print('Durbin-Watson Test')
dw_value = durbin_watson(residuals)
print("Durbin-Watson statistic: ", round(dw_value,3))
```

```
Durbin-Watson Test
Durbin-Watson statistic:  1.011
```

In [12]:
```python
from scipy import stats

#Test for cross-section dependency

#H0: No cross-section dependency
```

```
#H1: Cross-section dependency exists

print('Breusch-Pagan LM Test')
resid_df = residuals.unstack(level=0)
T = resid_df.shape[0]
N = resid_df.shape[1]

rho = resid_df.corr().values
upper_tri_idx = np.triu_indices(N, k=1)
rho_upper = rho[upper_tri_idx]
LM_stat = T * np.sum(rho_upper**2)
p_value = 1 - stats.chi2.cdf(LM_stat, N*(N-1)/2)

print(f"Breusch-Pagan LM statistic: {LM_stat:.3f}")
print(f"p-value: {p_value:.4f}")
print()

print('Pesaran CD Test')
CD_stat = np.sqrt(2 / (N*(N-1))) * np.sum(rho_upper)
p_value_cd = 2 * (1 - stats.norm.cdf(abs(CD_stat)))

print(f"Pesaran CD statistic: {CD_stat:.3f}")
print(f"p-value: {p_value_cd:.4f}")
```

```
Breusch-Pagan LM Test
Breusch-Pagan LM statistic: 1638.685
p-value: 0.0000

Pesaran CD Test
Pesaran CD statistic: 8.856
p-value: 0.0000
```

In [13]:
```
#Re-estimate RE Model
```

In [14]:
```
#RE with cov.type 'clustered'
re_model_robust1 = REmodel.fit(cov_type='clustered', cluster_entity=True)
print(re_model_robust1.summary)
```

```
                          RandomEffects Estimation Summary
================================================================================
Dep. Variable:             violent_cr   R-squared:                        0.0033
Estimator:              RandomEffects   R-squared (Between):              0.0549
No. Observations:                 384   R-squared (Within):              -0.0016
Date:                Wed, Nov 12 2025   R-squared (Overall):              0.0285
Time:                        18:28:44   Log-likelihood                   -1167.7
Cov. Estimator:             Clustered
                                        F-statistic:                      0.6212
Entities:                          32   P-value                           0.5379
Avg Obs:                       12.000   Distribution:                   F(2,381)
Min Obs:                       12.000
Max Obs:                       12.000   F-statistic (robust):             0.7032
                                        P-value                           0.4956
Time periods:                      12   Distribution:                   F(2,381)
Avg Obs:                       32.000
Min Obs:                       32.000
Max Obs:                       32.000

                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
const          8.9686     6.1474     1.4589     0.1454     -3.1186      21.056
avg_temp       0.2500     0.2123     1.1777     0.2397     -0.1674      0.6674
tot_rf         0.0005     0.0010     0.4719     0.6373     -0.0015      0.0025
==============================================================================
```

In [15]:

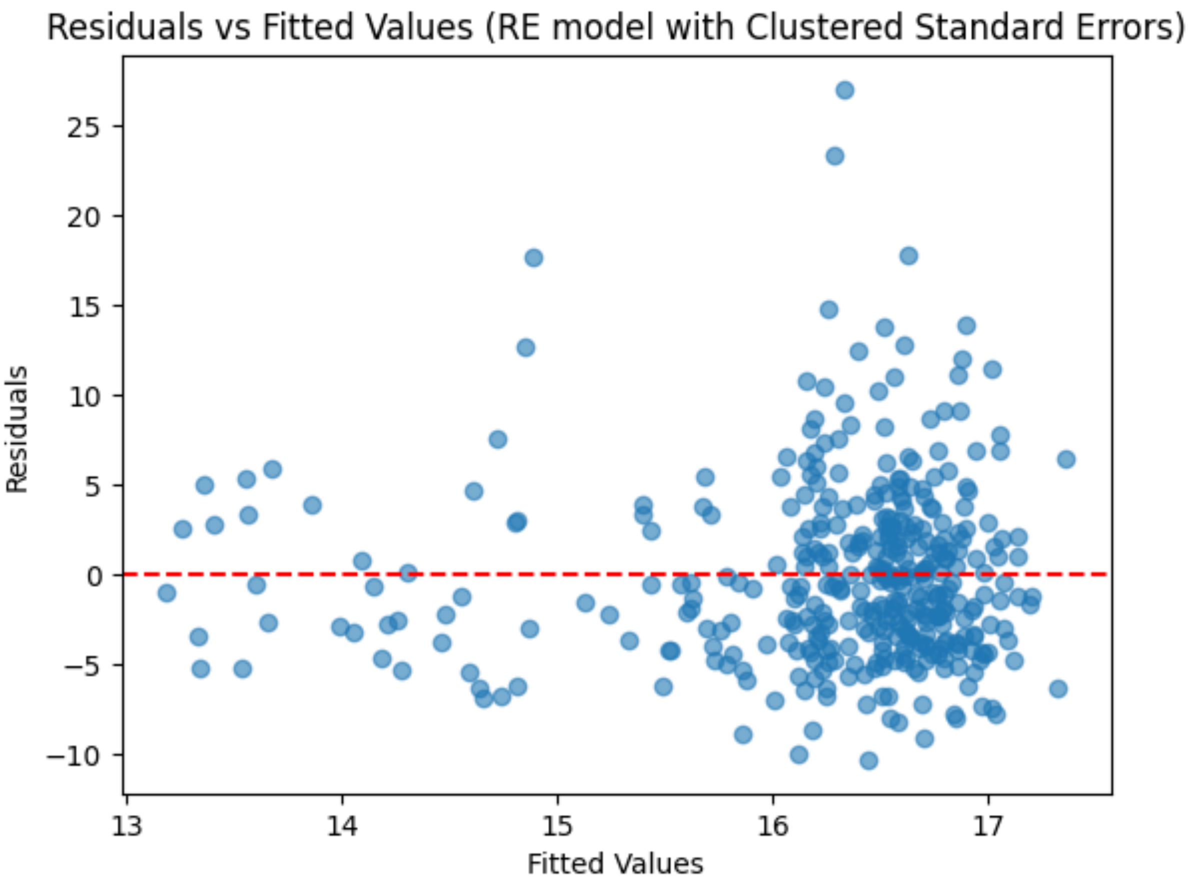
```python
#RE with cov.type 'kernel' (Driscoll-Kraay Method)
re_model_robust2 = REmodel.fit(cov_type='kernel')
print(re_model_robust2.summary)
```

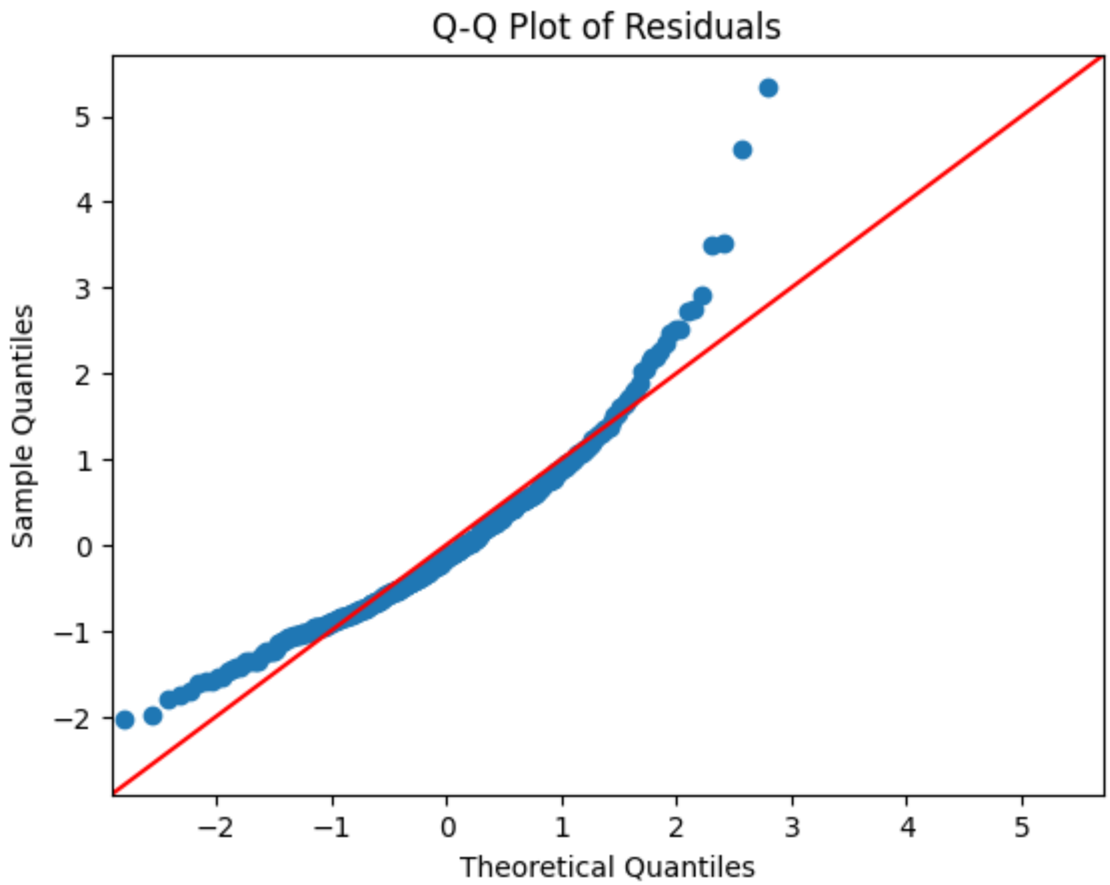
```
                          RandomEffects Estimation Summary
================================================================================
Dep. Variable:             violent_cr   R-squared:                        0.0033
Estimator:              RandomEffects   R-squared (Between):              0.0549
No. Observations:                 384   R-squared (Within):              -0.0016
Date:                Wed, Nov 12 2025   R-squared (Overall):              0.0285
Time:                        18:28:44   Log-likelihood                   -1167.7
Cov. Estimator:         Driscoll-Kraay
                                        F-statistic:                      0.6212
Entities:                          32   P-value                           0.5379
Avg Obs:                       12.000   Distribution:                    F(2,381)
Min Obs:                       12.000
Max Obs:                       12.000   F-statistic (robust):             1.0834
                                        P-value                           0.3395
Time periods:                      12   Distribution:                    F(2,381)
Avg Obs:                       32.000
Min Obs:                       32.000
Max Obs:                       32.000

                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
const          8.9686     8.3392     1.0755     0.2828     -7.4280      25.365
avg_temp       0.2500     0.2355     1.0615     0.2891     -0.2131      0.7131
tot_rf         0.0005     0.0010     0.4633     0.6434     -0.0016      0.0025
==============================================================================
```

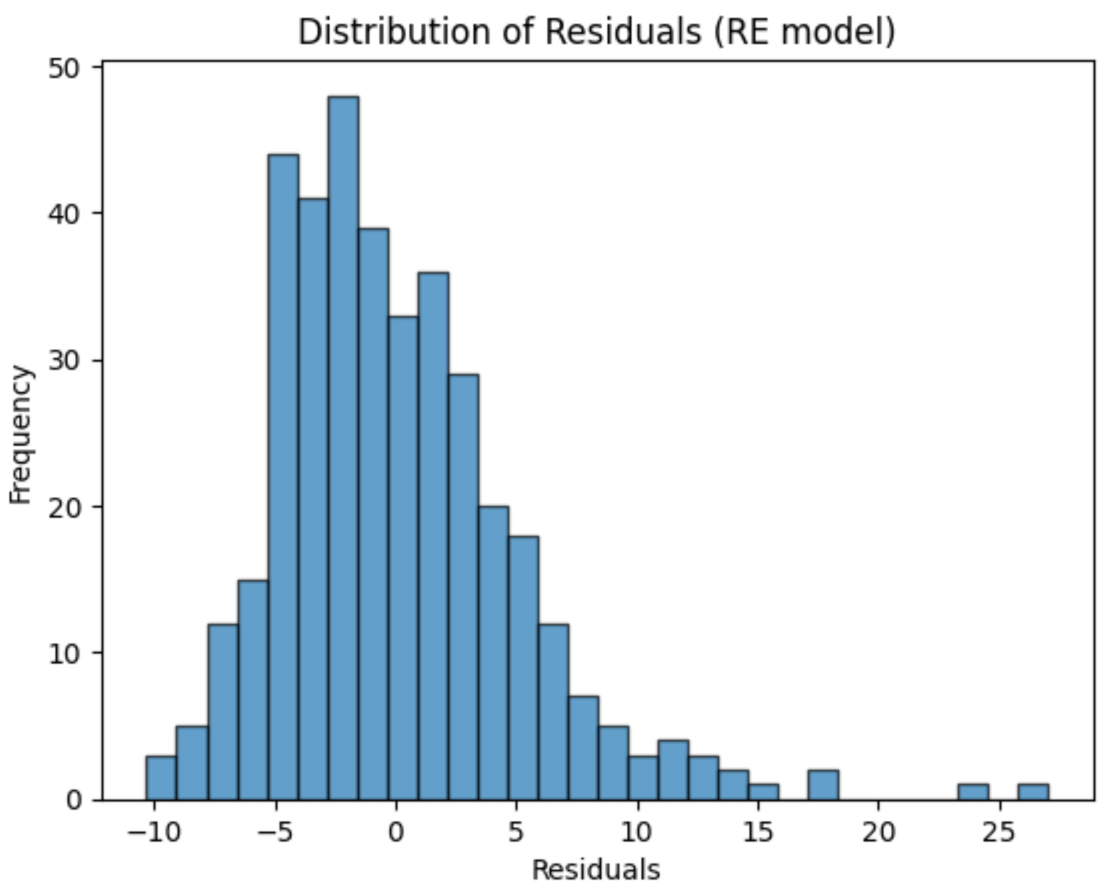
In [16]:

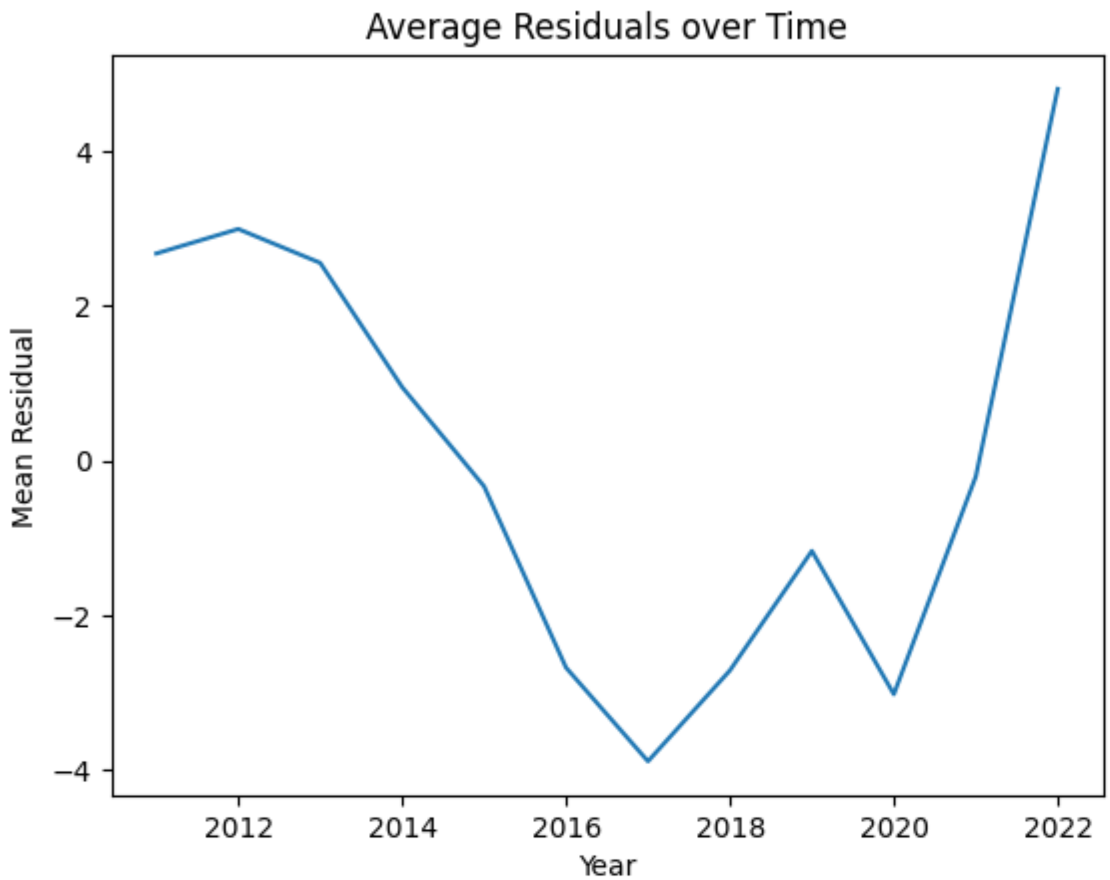
```python
# Check residuals and fitted values
df['residuals1'] = re_model_robust1.resids
df['fitted1'] = re_model_robust1.fitted_values

import matplotlib.pyplot as plt

plt.scatter(df['fitted1'], df['residuals1'], alpha=0.6)
plt.axhline(0, color='red', linestyle='--')
plt.xlabel('Fitted Values')
plt.ylabel('Residuals')
plt.title('Residuals vs Fitted Values (RE model with Clustered Standard Errors)')
plt.show()

sm.qqplot(df['residuals1'], line='45', fit=True)
plt.title('Q-Q Plot of Residuals')
plt.show()

plt.hist(df['residuals1'], bins=30, edgecolor='black', alpha=0.7)
plt.xlabel('Residuals')
plt.ylabel('Frequency')
plt.title('Distribution of Residuals (RE model)')
plt.show()

resid_df = df['residuals1'].unstack(level=0)
plt.plot(resid_df.mean(axis=1))
plt.title('Average Residuals over Time')
plt.xlabel('Year')
plt.ylabel('Mean Residual')
```

```
plt.show()

from scipy.stats import shapiro

#Test for normality

stat, p = shapiro(df['residuals1'])
print(f"Shapiro-Wilk Test: Statistic={stat:.3f}, p-value={p:.4f}")
```




Q-Q Plot of Residuals
Sample Quantiles
Theoretical Quantiles
−2
−1
0
1
2
3
4
5

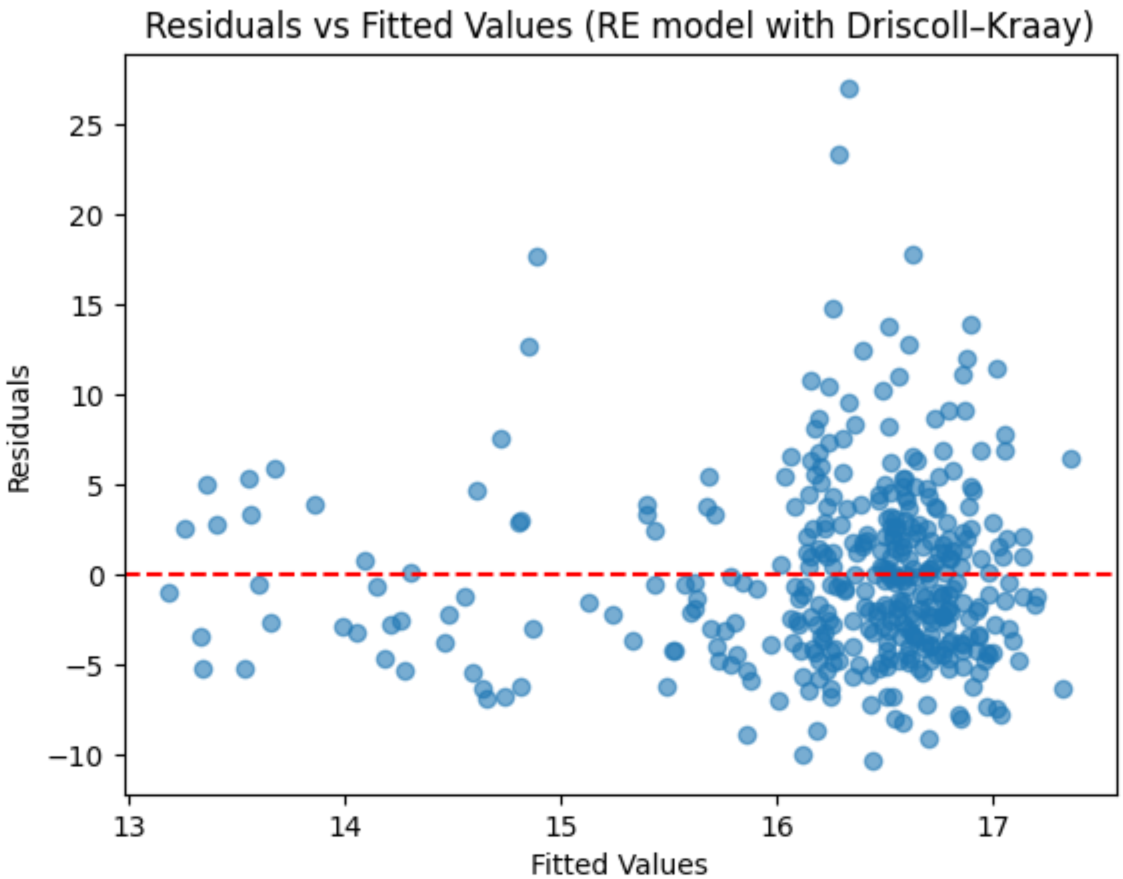

Distribution of Residuals (RE model)
Frequency
Residuals
0
10
20
30
40
50
−10
−5
0
5
10
15
20
25

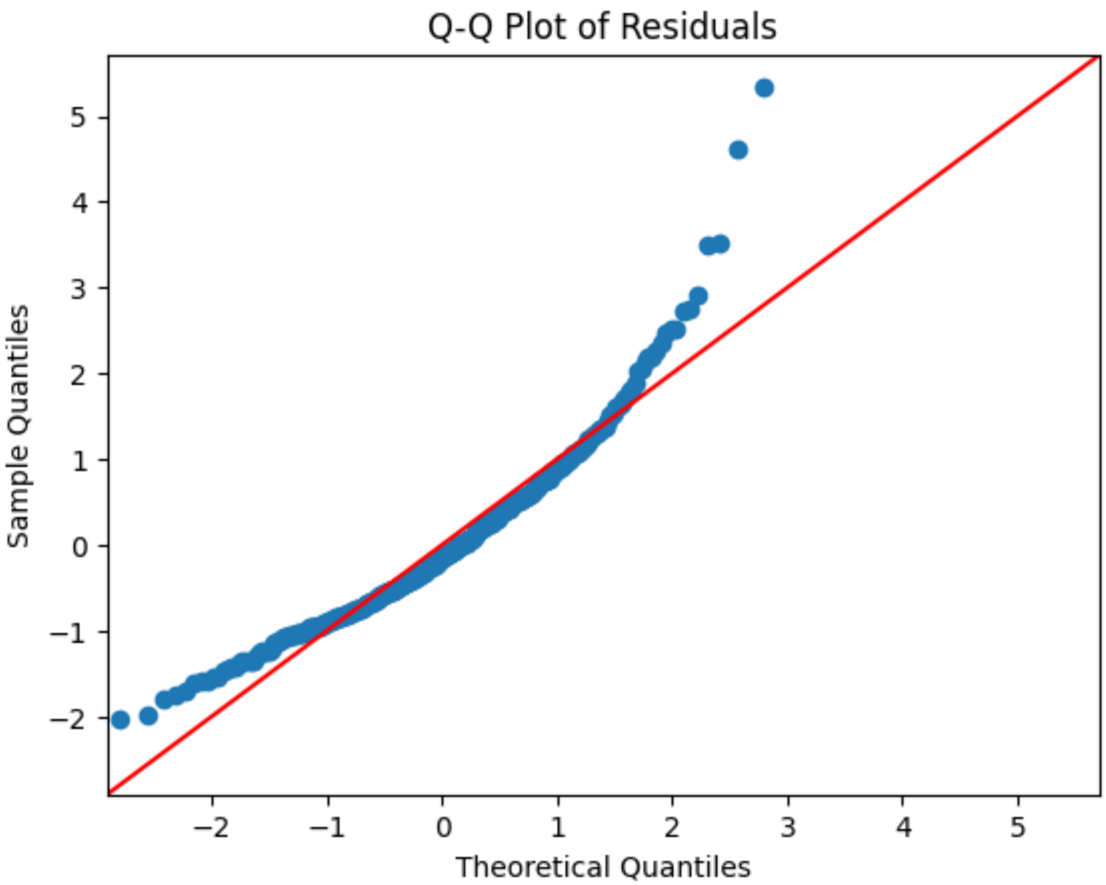


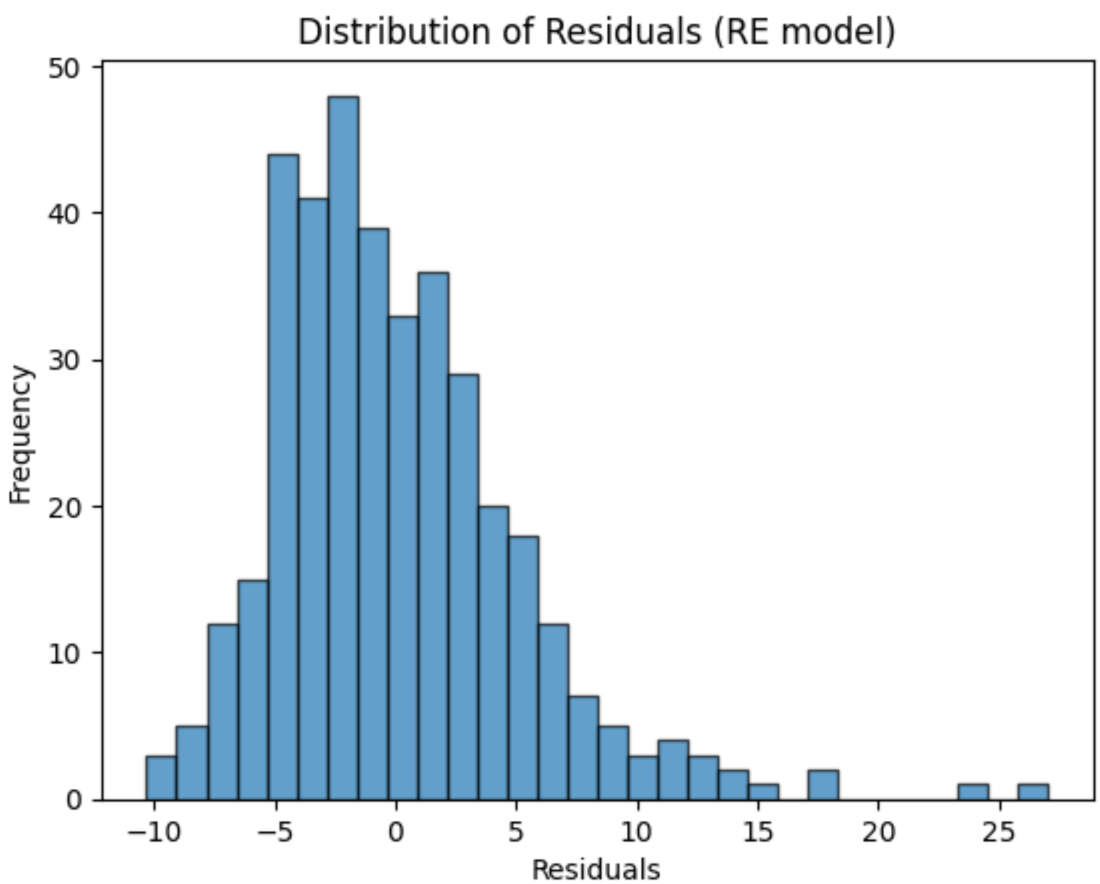

```
Shapiro-Wilk Test: Statistic=0.932, p-value=0.0000
```

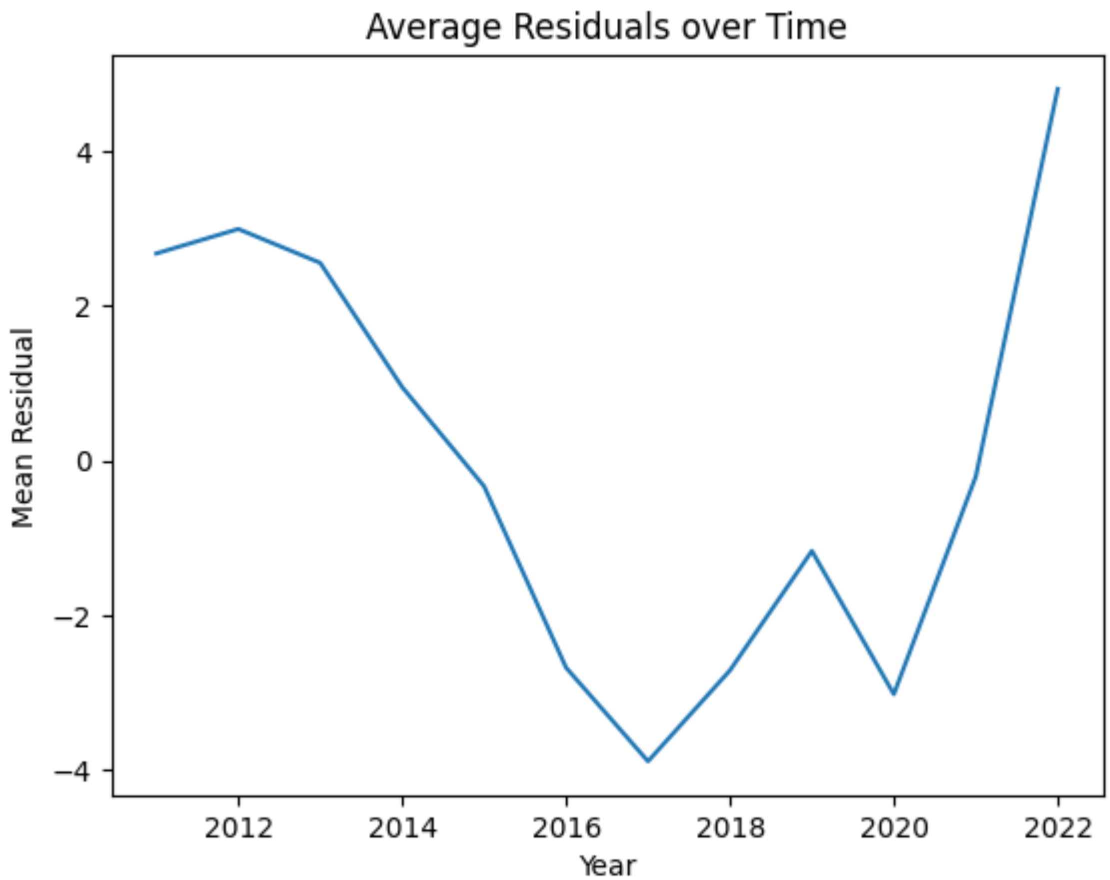
In [17]:
```
# Check residuals and fitted values
df['residuals2'] = re_model_robust2.resids
df['fitted2'] = re_model_robust2.fitted_values

import matplotlib.pyplot as plt

plt.scatter(df['fitted2'], df['residuals2'], alpha=0.6)
plt.axhline(0, color='red', linestyle='--')
plt.xlabel('Fitted Values')
plt.ylabel('Residuals')
plt.title('Residuals vs Fitted Values (RE model with Driscoll-Kraay)')
plt.show()

sm.qqplot(df['residuals2'], line='45', fit=True)
plt.title('Q-Q Plot of Residuals')
plt.show()

plt.hist(df['residuals2'], bins=30, edgecolor='black', alpha=0.7)
plt.xlabel('Residuals')
plt.ylabel('Frequency')
plt.title('Distribution of Residuals (RE model)')
plt.show()

resid_df = df['residuals2'].unstack(level=0)
plt.plot(resid_df.mean(axis=1))
plt.title('Average Residuals over Time')
plt.xlabel('Year')
```

```
plt.ylabel('Mean Residual')
plt.show()

from scipy.stats import shapiro

#Test for normality

stat, p = shapiro(df['residuals2'])
print(f"Shapiro-Wilk Test: Statistic={stat:.3f}, p-value={p:.4f}")
```




Q-Q Plot of Residuals
Sample Quantiles
Theoretical Quantiles
5
4
3
2
1
0
−1
−2
−2
−1
0
1
2
3
4
5


Distribution of Residuals (RE model)
Frequency
Residuals
50
40
30
20
10
0
−10
−5
0
5
10
15
20
25




```
Shapiro-Wilk Test: Statistic=0.932, p-value=0.0000
```

In [ ]: