import pandas as pd

from google.colab import files

# 파일 업로드 창 열기

uploaded = files.upload()

# 필요한 라이브러리 설치

!pip install openpyxl statsmodels

# 라이브러리 임포트

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import statsmodels.api as sm

from statsmodels.tsa.stattools import adfuller, coint

# 엑셀 파일 읽기

file\_path = 'data국가\_온실가스\_종류별\_배출량\_추이\_20250517175021.xlsx'

# 파일 위치

df = pd.read\_excel(file\_path, header=0)

# ADF 검정을 각각의 열에 대해 수행

for column in df.columns:

    print(f"\n{column}에 대한 ADF 검정 결과:")

    # ADF 검정 실행

    result = adfuller(df[column].dropna())  # 결측값 제거 후 실행

    # 결과 출력

    print("ADF Statistic: ", result[0])

    print("p-value: ", result[1])

    print("Critical Values: ", result[4])

    # 결론: p-value에 따라 단위근 존재 여부 판별

    if result[1] < 0.05:

        print(f"{column} 열은 정상성(Stationarity)을 가집니다.")

    else:

        print(f"{column} 열은 정상성이 없습니다.")

insurance = df['원수보험금']

CO2 = df['CO₂']

GDP = df['국내총생산(시장가격)']

bu = df['부적업소']

insurance\_diff = insurance.diff().dropna()

CO2\_diff = CO2.diff().dropna()

GDP\_diff = GDP.diff().dropna()

bu\_diff = bu.diff().dropna()

df.head()

X = df[['CO₂', '국내총생산(시장가격)', '부적업소']]

X = sm.add\_constant(X)

y = df['원수보험금']

# 선형 회귀 (Engle-Granger Step 1)

model = sm.OLS(y, X).fit()

print(model.summary())

# 잔차 계산

residuals = model.resid

# 잔차에 대해 단위근 검정 (ADF Test)

adf\_result = adfuller(residuals)

print(f"ADF Statistic: {adf\_result[0]}")

print(f"p-value: {adf\_result[1]}")

for key, value in adf\_result[4].items():

    print(f'Critical Value ({key}): {value}')

# 데이터프레임으로 정리

ecm\_df = pd.DataFrame({

    'CO2': CO2,

    'insurance': insurance,

    'GDP': GDP,

    'bu': bu,

    'residuals': residuals

})

# 1차 차분 (단기 변화량)

ecm\_df['d\_CO2'] = CO2\_diff

ecm\_df['d\_insurance'] = insurance\_diff

ecm\_df['d\_GDP'] = GDP\_diff

ecm\_df['d\_bu'] = bu\_diff

# 오차수정항 (잔차의 lag)

ecm\_df['ec\_term\_lag'] = ecm\_df['residuals'].shift(1)

# 결측치 제거

ecm\_df.dropna(inplace=True)

# 독립 변수 구성: 단기 변화 + 오차수정항

X\_ecm = sm.add\_constant(ecm\_df[['d\_CO2', 'd\_GDP', 'd\_bu', 'ec\_term\_lag' ]])

y\_ecm = ecm\_df['d\_insurance']

# ECM 회귀

ecm\_model = sm.OLS(y\_ecm, X\_ecm).fit()

print(ecm\_model.summary())

plt.figure(figsize=(12, 4))

plt.plot(ecm\_df['residuals'], label='out of long-run equilibrium (error term)')

plt.axhline(0, color='red', linestyle='--')

plt.title("time series trend of error term(out of long-run equilibrium)")

plt.ylabel("residuals (ε)")

plt.xlabel("time")

plt.legend()

plt.show()