In [1]:

```python
import pandas as pd
import numpy as np
from sklearn.linear_model import Ridge
from sklearn.metrics import mean_squared_error, mean_absolute_error
from sklearn.model_selection import KFold
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline
```

In [2]:

```python
# Load the dataset
data = pd.read_csv('1M_ahead_dataset.csv')
```

In [3]:

```python
# Assume that 'Yt.1M' is the target and the rest are predictors
X = data.drop(['Yt.1M'], axis=1)
y = data['Yt.1M']
```

In [4]:

```python
# Define candidate values for alpha (λ) to test.
alphas = [305, 306, 307, 308, 309]
```

In [5]:

```python
# Set up 5-Fold cross-validation
kf = KFold(n_splits=5, shuffle=True, random_state=42)
```

In [6]:

```python
# To store the average performance for each alpha candidate.
results = []
```

In [7]:

```python
# To store coefficients for each fold
coefficient_results = []
```

In [8]:

```python
# Loop over each candidate alpha
for alpha in alphas:
    fold_metrics = []
    print(f"\n--- Alpha {alpha} ---")
    fold_counter = 1
    for train_index, test_index in kf.split(X):
        # Split into training and testing sets for this fold
        X_train, X_test = X.iloc[train_index], X.iloc[test_index]
        y_train, y_test = y.iloc[train_index], y.iloc[test_index]

        # Create a pipeline that scales the data and then applies Ridge regression
        model = Pipeline([
            ('scaler', StandardScaler()),
            ('ridge', Ridge(alpha=alpha))
        ])

        # Fit the model on training data
        model.fit(X_train, y_train)

        # Extract coefficients and intercept from the Ridge model
        # (Note: These coefficients correspond to scaled features)
        coef = model.named_steps['ridge'].coef_
        intercept = model.named_steps['ridge'].intercept_

        print(f"\nFold {fold_counter} Coefficients:")
        print("Coefficients:", coef)
        print("Intercept:", intercept)

        # Save the coefficients info
        coefficient_results.append({
            'alpha': alpha,
            'fold': fold_counter,
            'coefficients': coef,
            'intercept': intercept
        })

        # Make predictions for the test set
        y_pred = model.predict(X_test)

        # Compute evaluation metrics
        mse = mean_squared_error(y_test, y_pred)
        mae = mean_absolute_error(y_test, y_pred)
        rmse = np.sqrt(mse)

        fold_metrics.append({'MSE': mse, 'RMSE': rmse, 'MAE': mae})

        # Increment fold counter
        fold_counter += 1

    # Compute average metrics over 5 folds for current alpha
    avg_mse = np.mean([m['MSE'] for m in fold_metrics])
    avg_rmse = np.mean([m['RMSE'] for m in fold_metrics])
    avg_mae = np.mean([m['MAE'] for m in fold_metrics])
    results.append({'alpha': alpha, 'avg_MSE': avg_mse, 'avg_RMSE': avg_rmse, 'avg_MAE': avg_mae})
    print(f"\nAlpha {alpha} -- Avg MSE: {avg_mse:.4f}, Avg RMSE: {avg_rmse:.4f}, Avg MAE: {avg_mae:.4f}")
```

```
--- Alpha 305 ---

Fold 1 Coefficients:
Coefficients: [ 0.0112779   0.00093962  0.00136772  0.01437968 -0.00086332  0.00807467
 -0.00037825 -0.00395971 -0.00531684  0.00616475 -0.00403575  0.00904481
  0.00059337]
Intercept: 0.04288559670781893

Fold 2 Coefficients:
Coefficients: [ 0.0103322   0.00015042  0.00231984  0.01553257 -0.00331228  0.00598527
 -0.00053489 -0.00323713 -0.00436739  0.00518027 -0.0013217   0.01092383
  0.00067608]
Intercept: 0.044819341563786

Fold 3 Coefficients:
Coefficients: [ 0.00672094  0.00164141  0.00278518  0.01492215 -0.00065732  0.00590635
 -0.000569   -0.00499564 -0.00884486  0.00665053 -0.00124285  0.00560227
  0.00311413]
Intercept: 0.0430213698630137

Fold 4 Coefficients:
Coefficients: [ 0.00626842  0.00261658  0.00316873  0.01559258 -0.00117358  0.00930485
 -0.00056288 -0.00352702 -0.00881027  0.00544224 -0.00405974  0.00719789
  0.00245405]
Intercept: 0.04924054794520548

Fold 5 Coefficients:
Coefficients: [ 8.81535025e-03  4.02550616e-03  2.59285965e-03  1.70487439e-02
 -7.05436424e-04  5.39575640e-03  3.45429370e-05 -2.93855546e-03
 -9.07738401e-03  6.35666389e-03 -2.85887959e-03  9.64138675e-03
  2.13041167e-03]
Intercept: 0.04469027397260275

Alpha 305 -- Avg MSE: 0.0154, Avg RMSE: 0.1235, Avg MAE: 0.0728

--- Alpha 306 ---

Fold 1 Coefficients:
Coefficients: [ 0.01126527  0.00094307  0.00136623  0.01436736 -0.00086125  0.00806736
 -0.00037808 -0.00395621 -0.00530667  0.00616025 -0.00403123  0.00903558
  0.00059421]
Intercept: 0.04288559670781893

Fold 2 Coefficients:
Coefficients: [ 0.01032076  0.00015674  0.0023158   0.01551802 -0.00330701  0.0059808
 -0.00053465 -0.00323414 -0.00435935  0.00517551 -0.00132091  0.01091288
  0.00067752]
Intercept: 0.044819341563786

Fold 3 Coefficients:
Coefficients: [ 0.00671415  0.00164292  0.00278068  0.01490788 -0.00065548  0.00590198
 -0.00056881 -0.00499307 -0.00883115  0.00664469 -0.00124115  0.00559491
  0.00311182]
Intercept: 0.0430213698630137

Fold 4 Coefficients:
Coefficients: [ 0.00626104  0.0026179   0.00316545  0.01557896 -0.0011713   0.0092966
 -0.00056256 -0.00352715 -0.00879601  0.00543873 -0.00405474  0.00719061
  0.00245306]
Intercept: 0.04924054794520548

Fold 5 Coefficients:
Coefficients: [ 8.80650298e-03  4.02322032e-03  2.59230475e-03  1.70334528e-02
 -7.03498478e-04  5.39114549e-03  4.12700740e-05 -2.93928145e-03
 -9.06167650e-03  6.35177744e-03 -2.85530018e-03  9.62921460e-03
  2.12932500e-03]
Intercept: 0.04469027397260275

Alpha 306 -- Avg MSE: 0.0154, Avg RMSE: 0.1235, Avg MAE: 0.0728

--- Alpha 307 ---

Fold 1 Coefficients:
Coefficients: [ 0.01125268  0.0009465   0.00136474  0.01435507 -0.00085919  0.00806007
 -0.00037791 -0.00395271 -0.00529654  0.00615575 -0.00402672  0.00902638
  0.00059504]
Intercept: 0.04288559670781893

Fold 2 Coefficients:
Coefficients: [ 0.01030935  0.00016302  0.00231179  0.01550349 -0.00330176  0.00597633
```

```
 -0.00053442 -0.00323116 -0.00435134  0.00517075 -0.00132011  0.01090194
  0.00067895]
Intercept: 0.044819341563786

Fold 3 Coefficients:
Coefficients: [ 0.00670737  0.0016444   0.00277619  0.01489363 -0.00065365  0.00589761
 -0.00056863 -0.00499049 -0.00881748  0.00663885 -0.00123945  0.00558757
  0.00310952]
Intercept: 0.0430213698630137

Fold 4 Coefficients:
Coefficients: [ 0.00625368  0.0026192   0.00316218  0.01556535 -0.00116903  0.00928836
 -0.00056224 -0.00352727 -0.0087818   0.00543522 -0.00404976  0.00718334
  0.00245208]
Intercept: 0.04924054794520548

Fold 5 Coefficients:
Coefficients: [ 8.79767713e-03  4.02093714e-03  2.59174711e-03  1.70181880e-02
 -7.01565878e-04  5.38654775e-03  4.79644501e-05 -2.93998579e-03
 -9.04602747e-03  6.34689763e-03 -2.85173242e-03  9.61708172e-03
  2.12824012e-03]
Intercept: 0.04469027397260275

Alpha 307 -- Avg MSE: 0.0154, Avg RMSE: 0.1235, Avg MAE: 0.0728

--- Alpha 308 ---

Fold 1 Coefficients:
Coefficients: [ 0.01124011  0.00094991  0.00136327  0.01434279 -0.00085713  0.0080528
 -0.00037773 -0.00394921 -0.00528644  0.00615126 -0.00402222  0.00901719
  0.00059587]
Intercept: 0.04288559670781893

Fold 2 Coefficients:
Coefficients: [ 0.01029797  0.00016925  0.0023078   0.01548899 -0.00329652  0.00597187
 -0.00053418 -0.00322817 -0.00434336  0.005166   -0.00131932  0.01089103
  0.00068038]
Intercept: 0.044819341563786

Fold 3 Coefficients:
Coefficients: [ 0.00670062  0.00164586  0.00277173  0.01487942 -0.00065182  0.00589326
 -0.00056844 -0.0049879  -0.00880386  0.00663303 -0.00123777  0.00558025
  0.00310722]
Intercept: 0.0430213698630137

Fold 4 Coefficients:
Coefficients: [ 0.00624634  0.00262048  0.00315893  0.01555178 -0.00116676  0.00928014
 -0.00056192 -0.00352737 -0.00876764  0.00543172 -0.0040448   0.0071761
  0.00245109]
Intercept: 0.04924054794520548

Fold 5 Coefficients:
Coefficients: [ 8.78887261e-03  4.01865661e-03  2.59118676e-03  1.70029494e-02
 -6.99638608e-04  5.38196308e-03  5.46262614e-05 -2.94066869e-03
 -9.03043657e-03  6.34202446e-03 -2.84817624e-03  9.60498791e-03
  2.12715703e-03]
Intercept: 0.04469027397260275

Alpha 308 -- Avg MSE: 0.0154, Avg RMSE: 0.1235, Avg MAE: 0.0728

--- Alpha 309 ---

Fold 1 Coefficients:
Coefficients: [ 0.01122758  0.00095328  0.00136181  0.01433054 -0.00085507  0.00804554
 -0.00037756 -0.00394571 -0.00527638  0.00614677 -0.00401773  0.00900803
  0.00059669]
Intercept: 0.04288559670781893

Fold 2 Coefficients:
Coefficients: [ 0.01028662  0.00017544  0.00230385  0.01547452 -0.00329129  0.00596742
 -0.00053394 -0.00322518 -0.0043354   0.00516126 -0.00131852  0.01088014
  0.0006818 ]
Intercept: 0.044819341563786

Fold 3 Coefficients:
Coefficients: [ 0.00669388  0.0016473   0.00276729  0.01486523 -0.00065     0.00588891
 -0.00056825 -0.0049853  -0.00879028  0.00662722 -0.00123608  0.00557295
  0.00310492]
Intercept: 0.0430213698630137

Fold 4 Coefficients:
Coefficients: [ 0.00623903  0.00262174  0.00315569  0.01553822 -0.00116449  0.00927193
 -0.0005616  -0.00352745 -0.00875353  0.00542822 -0.00403985  0.00716886
```

```
  0.0024501 ]
Intercept: 0.04924054794520548

Fold 5 Coefficients:
Coefficients: [ 8.78008933e-03  4.01637874e-03  2.59062373e-03  1.69877371e-02
 -6.97716650e-04  5.37739141e-03  6.12557025e-05 -2.94133035e-03
 -9.01490344e-03  6.33715791e-03 -2.84463156e-03  9.59293294e-03
  2.12607570e-03]
Intercept: 0.04469027397260275

Alpha 309 -- Avg MSE: 0.0154, Avg RMSE: 0.1235, Avg MAE: 0.0728
```

In [9]:

```
# Summarize the overall results in a DataFrame
results_df = pd.DataFrame(results)
print("\nOverall Cross-Validation Results:")
print(results_df)
```

```
Overall Cross-Validation Results:
   alpha   avg_MSE  avg_RMSE   avg_MAE
0    305  0.015383  0.123498  0.072762
1    306  0.015382  0.123498  0.072767
2    307  0.015382  0.123498  0.072773
3    308  0.015382  0.123498  0.072778
4    309  0.015382  0.123498  0.072783
```

In [10]:

```
# Identify the best alpha based on average MSE
best_alpha = results_df.loc[results_df['avg_MSE'].idxmin()]['alpha']
print("\nBest alpha based on average MSE:", best_alpha)
```

```
Best alpha based on average MSE: 308.0
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js