

```
In [7]: import pandas as pd
import numpy as np
from lifelines import CoxPHFitter
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler, OneHotEncoder
import matplotlib.pyplot as plt

# Load the data from the .xlsx file
data = pd.read_excel('data1.xlsx')

# Define categorical variables
categorical_cols = ['SEX', 'CompositeStage', 'LNInvolment', 'Comorbidity', 'FamiliyHistoryOfCancer']
data[categorical_cols] = data[categorical_cols].astype('category')

# One-hot encode categorical variables
data_encoded = pd.get_dummies(data, columns=categorical_cols, drop_first=True)

# Standardize the covariates
scaler = StandardScaler()
data_encoded[['DEATH', 'AGE']] = scaler.fit_transform(data_encoded[['DEATH', 'AGE']])
```

```
In [9]: buckley_james_data = data_encoded[['Months', 'DEATH', 'AGE']] + [col for col in data_encoded.columns if col.startswith('SEX_')]

# Fit the Buckley-James model with custom options
cph = CoxPHFitter(penalizer=0.1)
cph.fit(buckley_james_data, 'Months', 'DEATH', show_progress=True)
print(cph.summary)
cph.plot()
plt.xlabel('Time')
plt.ylabel('Survival Probability')
plt.title('James Buckley Estimator')
plt.show()
```

Iteration 1: norm\_delta = 0.68300, step\_size = 0.9500, log\_lik = -1663.17959, newton\_decrement = 54.49713, seconds\_since\_start = 0.0  
 Iteration 2: norm\_delta = 0.08068, step\_size = 0.9500, log\_lik = -1616.75385, newton\_decrement = 1.16193, seconds\_since\_start = 0.1  
 Iteration 3: norm\_delta = 0.00455, step\_size = 0.9500, log\_lik = -1615.59252, newton\_decrement = 0.00340, seconds\_since\_start = 0.1  
 Iteration 4: norm\_delta = 0.00000, step\_size = 1.0000, log\_lik = -1615.58912, newton\_decrement = 0.00000, seconds\_since\_start = 0.1  
 Convergence success after 4 iterations.

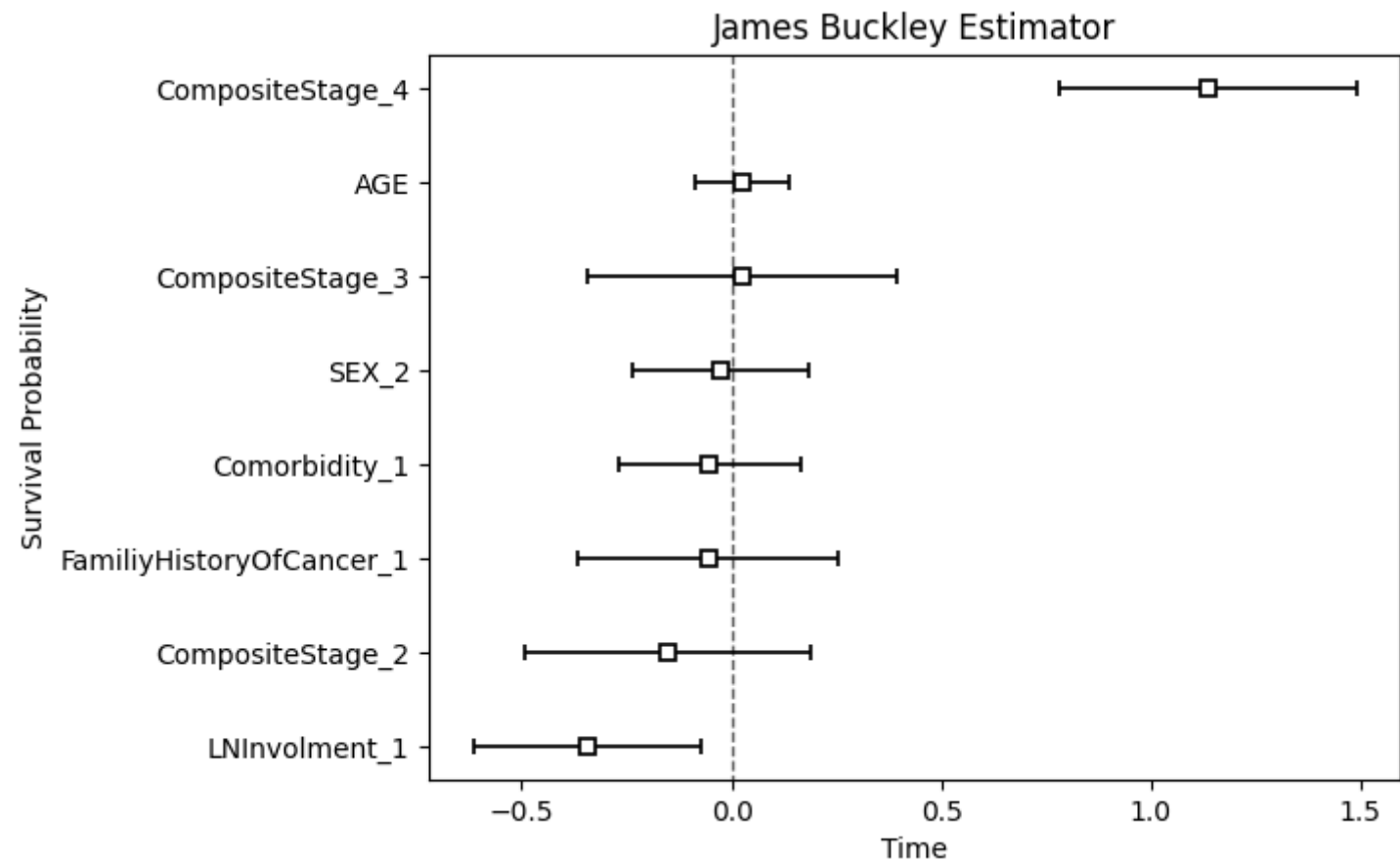
	coef	exp(coef)	se(coef)	coef lower 95% \
covariate				
AGE	0.026478	1.026831	0.056826	-0.084899
SEX_2	-0.027535	0.972841	0.107577	-0.238382
CompositeStage_2	-0.151530	0.859392	0.173728	-0.492030
CompositeStage_3	0.026319	1.026669	0.188558	-0.343248
CompositeStage_4	1.135923	3.114047	0.180333	0.782478
LNInvolment_1	-0.343457	0.709314	0.138901	-0.615698
Comorbidity_1	-0.053072	0.948312	0.111065	-0.270755
FamilyHistoryOfCancer_1	-0.055868	0.945664	0.157606	-0.364770

	coef upper 95%	exp(coef) lower 95% \
covariate		
AGE	0.137854	0.918605
SEX_2	0.183313	0.787902
CompositeStage_2	0.188969	0.611384
CompositeStage_3	0.395887	0.709462
CompositeStage_4	1.489369	2.186884
LNInvolment_1	-0.071217	0.540264
Comorbidity_1	0.164612	0.762803
FamilyHistoryOfCancer_1	0.253035	0.694356

	exp(coef) upper 95%	cmp to	z	p \
covariate				
AGE	1.147808	0.0	0.465942	6.412567e-01
SEX_2	1.201190	0.0	-0.255952	7.979879e-01
CompositeStage_2	1.208004	0.0	-0.872229	3.830834e-01
CompositeStage_3	1.485701	0.0	0.139582	8.889905e-01
CompositeStage_4	4.434295	0.0	6.299046	2.994829e-10
LNInvolment_1	0.931260	0.0	-2.472683	1.341029e-02
Comorbidity_1	1.178936	0.0	-0.477843	6.327621e-01

FamilyHistoryOfCancer\_1 1.287929 0.0 -0.354475 7.229829e-01

	-log2(p)
covariate	
AGE	0.641026
SEX_2	0.325561
CompositeStage_2	1.384270
CompositeStage_3	0.169760
CompositeStage_4	31.636807
LNInvolment_1	6.220515
Comorbidity_1	0.660265
FamilyHistoryOfCancer_1	0.467966



```
In [11]: n = len(buckley_james_data)
llf = cph.log_likelihood_
k = cph.params_.shape[0]
aic = -2 * llf + 2 * k
bic = -2 * llf + k * np.log(n)
# Print AIC and BIC
print("AIC value of the above data:", aic)
print("BIC value of the above data:", bic)
```

AIC value of the above data: 3247.1782463546083  
BIC value of the above data: 3277.880089931936

```
In [4]: import pandas as pd
import numpy as np
from lifelines import CoxPHFitter
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Load the data from the .xlsx file
data = pd.read_excel('data1.xlsx')

# Define categorical variables
categorical_cols = ['SEX', 'CompositeStage', 'LNInvolment', 'Comorbidity', 'FamiliyHistoryOfCancer']
data[categorical_cols] = data[categorical_cols].astype('category')

# One-hot encode categorical variables
data_encoded = pd.get_dummies(data, columns=categorical_cols, drop_first=True)

# Standardize the covariates
scaler = StandardScaler()
data_encoded[['DEATH', 'AGE']] = scaler.fit_transform(data_encoded[['DEATH', 'AGE']])
buckley_james_data = data_encoded[['Months', 'DEATH', 'AGE']] + [col for col in data_encoded.columns if col.startswith('SEX_')]

# Perform univariate analysis
univariate_results = []
for col in buckley_james_data.columns:
    if col not in ['Months', 'DEATH', 'AGE']:
        cph_univariate = CoxPHFitter(penalizer=0.1)
        cph_univariate.fit(buckley_james_data[[col, 'Months', 'DEATH', 'AGE']], 'Months', 'DEATH', show_progress=True)
        p_value = cph_univariate.summary['p'][col]
```

```

    univariate_results.append((col, p_value))
    n = len(buckley_james_data)
    llf = cph_univariate.log_likelihood_
    k = cph_univariate.params_.shape[0]
    aic = -2 * llf + 2 * k
    bic = -2 * llf + k * np.log(n)
    print(f"AIC value of {col}:", aic)
    print(f"BIC value of {col}:", bic)

# Select significant variables
significant_variables = [var for var, p_value in univariate_results if p_value < 0.05]
print("Significant variables from univariate analysis:", significant_variables)

# Fit the multivariate model
cph_multivariate = CoxPHFitter(penalizer=0.1)
cph_multivariate.fit(buckley_james_data[['Months', 'DEATH', 'AGE'] + significant_variables], 'Months', 'DEATH', show_progress=

# Calculate AIC and BIC
n = len(buckley_james_data)
llf = cph_multivariate.log_likelihood_
k = cph_multivariate.params_.shape[0]
aic = -2 * llf + 2 * k
bic = -2 * llf + k * np.log(n)

# Print AIC and BIC
print("AIC value of the multivariate model:", aic)
print("BIC value of the multivariate model:", bic)

# Plot the survival probability
cph_multivariate.plot()
plt.xlabel('Time')
plt.ylabel('Survival Probability')
plt.title('James Buckley Estimator')
plt.show()

```

Iteration 1: norm\_delta = 0.02554, step\_size = 0.9500, log\_lik = -1663.17959, newton\_decrement = 0.12225, seconds\_since\_start = 0.0  
Iteration 2: norm\_delta = 0.00120, step\_size = 0.9500, log\_lik = -1663.05786, newton\_decrement = 0.00027, seconds\_since\_start = 0.0  
Iteration 3: norm\_delta = 0.00006, step\_size = 0.9500, log\_lik = -1663.05758, newton\_decrement = 0.00000, seconds\_since\_start = 0.0  
Iteration 4: norm\_delta = 0.00000, step\_size = 1.0000, log\_lik = -1663.05758, newton\_decrement = 0.00000, seconds\_since\_start = 0.0  
Convergence success after 4 iterations.  
AIC value of SEX\_2: 3330.115165027162  
BIC value of SEX\_2: 3337.790625921494  
Iteration 1: norm\_delta = 0.14622, step\_size = 0.9500, log\_lik = -1663.17959, newton\_decrement = 4.31318, seconds\_since\_start = 0.0  
Iteration 2: norm\_delta = 0.01327, step\_size = 0.9500, log\_lik = -1658.78794, newton\_decrement = 0.03279, seconds\_since\_start = 0.0  
Iteration 3: norm\_delta = 0.00073, step\_size = 0.9500, log\_lik = -1658.75515, newton\_decrement = 0.00010, seconds\_since\_start = 0.0  
Iteration 4: norm\_delta = 0.00000, step\_size = 1.0000, log\_lik = -1658.75505, newton\_decrement = 0.00000, seconds\_since\_start = 0.0  
Convergence success after 4 iterations.  
AIC value of CompositeStage\_2: 3321.5101024525675  
BIC value of CompositeStage\_2: 3329.1855633468995  
Iteration 1: norm\_delta = 0.17583, step\_size = 0.9500, log\_lik = -1663.17959, newton\_decrement = 6.44524, seconds\_since\_start = 0.0  
Iteration 2: norm\_delta = 0.01738, step\_size = 0.9500, log\_lik = -1656.59487, newton\_decrement = 0.05734, seconds\_since\_start = 0.0  
Iteration 3: norm\_delta = 0.00098, step\_size = 0.9500, log\_lik = -1656.53746, newton\_decrement = 0.00018, seconds\_since\_start = 0.0  
Iteration 4: norm\_delta = 0.00000, step\_size = 1.0000, log\_lik = -1656.53728, newton\_decrement = 0.00000, seconds\_since\_start = 0.0  
Convergence success after 4 iterations.  
AIC value of CompositeStage\_3: 3317.0745678980034  
BIC value of CompositeStage\_3: 3324.7500287923353  
Iteration 1: norm\_delta = 0.70165, step\_size = 0.9500, log\_lik = -1663.17959, newton\_decrement = 50.56989, seconds\_since\_start = 0.0  
Iteration 2: norm\_delta = 0.08686, step\_size = 0.9500, log\_lik = -1620.40900, newton\_decrement = 1.08200, seconds\_since\_start = 0.0  
Iteration 3: norm\_delta = 0.00480, step\_size = 0.9500, log\_lik = -1619.32672, newton\_decrement = 0.00327, seconds\_since\_start = 0.0  
Iteration 4: norm\_delta = 0.00000, step\_size = 1.0000, log\_lik = -1619.32345, newton\_decrement = 0.00000, seconds\_since\_start =

0.0

Convergence success after 4 iterations.

AIC value of CompositeStage\_4: 3242.6468994706224

BIC value of CompositeStage\_4: 3250.3223603649544

Iteration 1: norm\_delta = 0.14048, step\_size = 0.9500, log\_lik = -1663.17959, newton\_decrement = 4.00489, seconds\_since\_start = 0.0

Iteration 2: norm\_delta = 0.01332, step\_size = 0.9500, log\_lik = -1659.09305, newton\_decrement = 0.03361, seconds\_since\_start = 0.0

Iteration 3: norm\_delta = 0.00074, step\_size = 0.9500, log\_lik = -1659.05942, newton\_decrement = 0.00010, seconds\_since\_start = 0.0

Iteration 4: norm\_delta = 0.00000, step\_size = 1.0000, log\_lik = -1659.05932, newton\_decrement = 0.00000, seconds\_since\_start = 0.0

Convergence success after 4 iterations.

AIC value of LNInvolment\_1: 3322.1186323122874

BIC value of LNInvolment\_1: 3329.7940932066194

Iteration 1: norm\_delta = 0.06728, step\_size = 0.9500, log\_lik = -1663.17959, newton\_decrement = 0.79933, seconds\_since\_start = 0.0

Iteration 2: norm\_delta = 0.00280, step\_size = 0.9500, log\_lik = -1662.38626, newton\_decrement = 0.00141, seconds\_since\_start = 0.0

Iteration 3: norm\_delta = 0.00014, step\_size = 0.9500, log\_lik = -1662.38485, newton\_decrement = 0.00000, seconds\_since\_start = 0.0

Iteration 4: norm\_delta = 0.00000, step\_size = 1.0000, log\_lik = -1662.38484, newton\_decrement = 0.00000, seconds\_since\_start = 0.0

Convergence success after 4 iterations.

AIC value of Comorbidity\_1: 3328.7696858848362

BIC value of Comorbidity\_1: 3336.445146779168

Iteration 1: norm\_delta = 0.03241, step\_size = 0.9500, log\_lik = -1663.17959, newton\_decrement = 0.19925, seconds\_since\_start = 0.0

Iteration 2: norm\_delta = 0.00091, step\_size = 0.9500, log\_lik = -1662.98363, newton\_decrement = 0.00017, seconds\_since\_start = 0.0

Iteration 3: norm\_delta = 0.00005, step\_size = 0.9500, log\_lik = -1662.98346, newton\_decrement = 0.00000, seconds\_since\_start = 0.0

Iteration 4: norm\_delta = 0.00000, step\_size = 1.0000, log\_lik = -1662.98346, newton\_decrement = 0.00000, seconds\_since\_start = 0.0

Convergence success after 4 iterations.

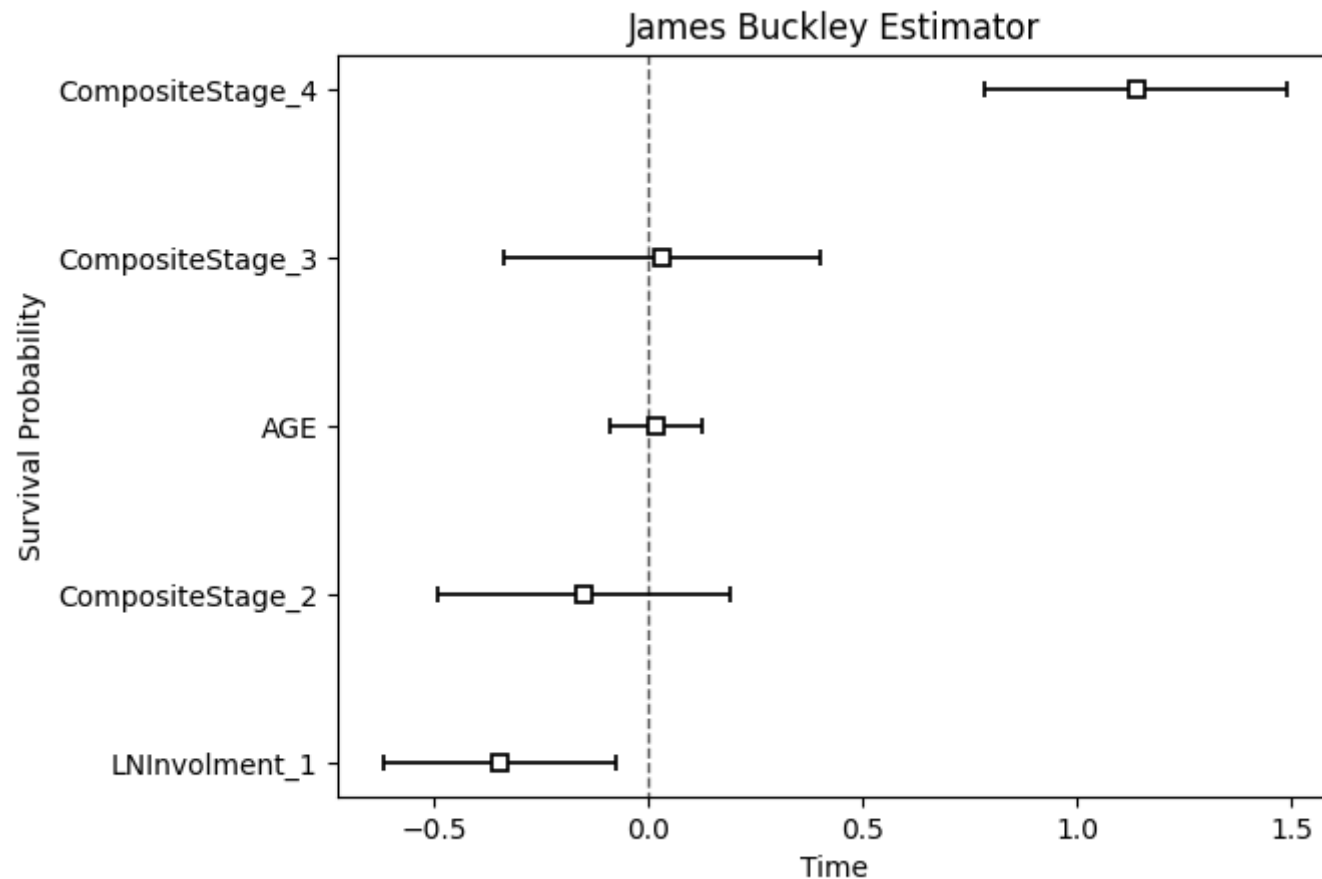
AIC value of FamilyHistoryOfCancer\_1: 3329.9669227876907

BIC value of FamilyHistoryOfCancer\_1: 3337.6423836820227

Significant variables from univariate analysis: ['CompositeStage\_2', 'CompositeStage\_3', 'CompositeStage\_4', 'LNInvolment\_1']

Iteration 1: norm\_delta = 0.68301, step\_size = 0.9500, log\_lik = -1663.17959, newton\_decrement = 54.30464, seconds\_since\_start = 0.0

```
Iteration 2: norm_delta = 0.08057, step_size = 0.9500, log_lik = -1616.92886, newton_decrement = 1.14381, seconds_since_start = 0.0
Iteration 3: norm_delta = 0.00450, step_size = 0.9500, log_lik = -1615.78591, newton_decrement = 0.00329, seconds_since_start = 0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1615.78262, newton_decrement = 0.00000, seconds_since_start = 0.0
Convergence success after 4 iterations.
AIC value of the multivariate model: 3241.5652333399553
BIC value of the multivariate model: 3260.753885575785
```



In [ ]: