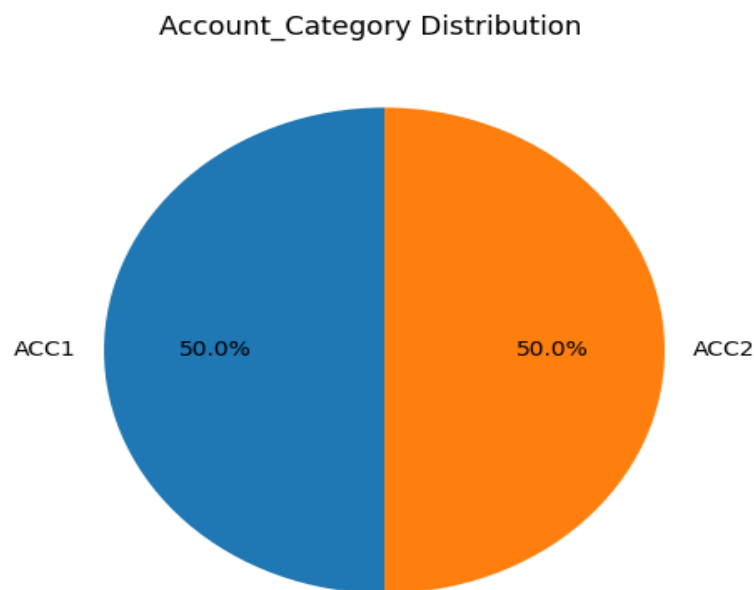


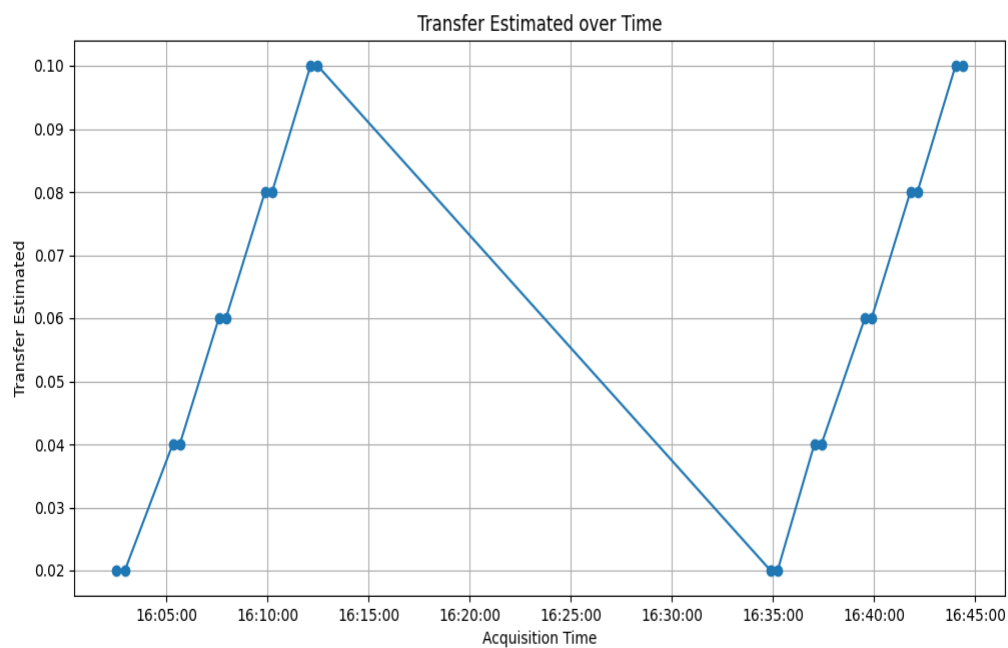
多模型报告1

关键信息分布可视化：

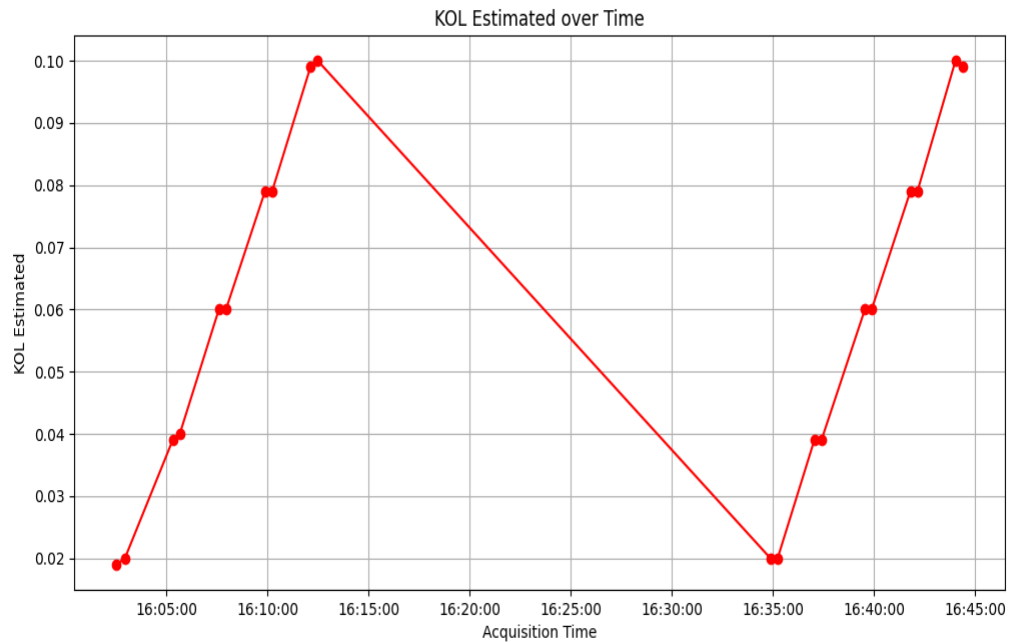
Account Category 分布饼图：



Transfer_Estimated 时序图：



KOL_Estimated 时序图:



数据处理:

1. 缺失值处理:

删去缺失值列 Note

2. 数据整理:

2.1 新建两列分别用列表储存x, Account_Corrected列的数据

2.2 对账户类型数据进行编码

3. 确定特征向量与目标向量:

3.1 特征向量:

每个向量由46个特征组成: x (22)+Account_Corrected (22)+编码化账户类型 (2)

3.2 目标向量:

每个向量由2个特征组成: KOL_Estimated (1)+Transfer_Estimated (1)

4. 数据集划分:

训练集 (16) + 测试集 (4)

训练模型：

1. 线性回归模型

1.1 参数

原始模型参数：

```
{'copy_X': True, 'fit_intercept': True, 'n_jobs': None, 'normalize':  
'deprecated', 'positive': False}
```

调参后模型参数：

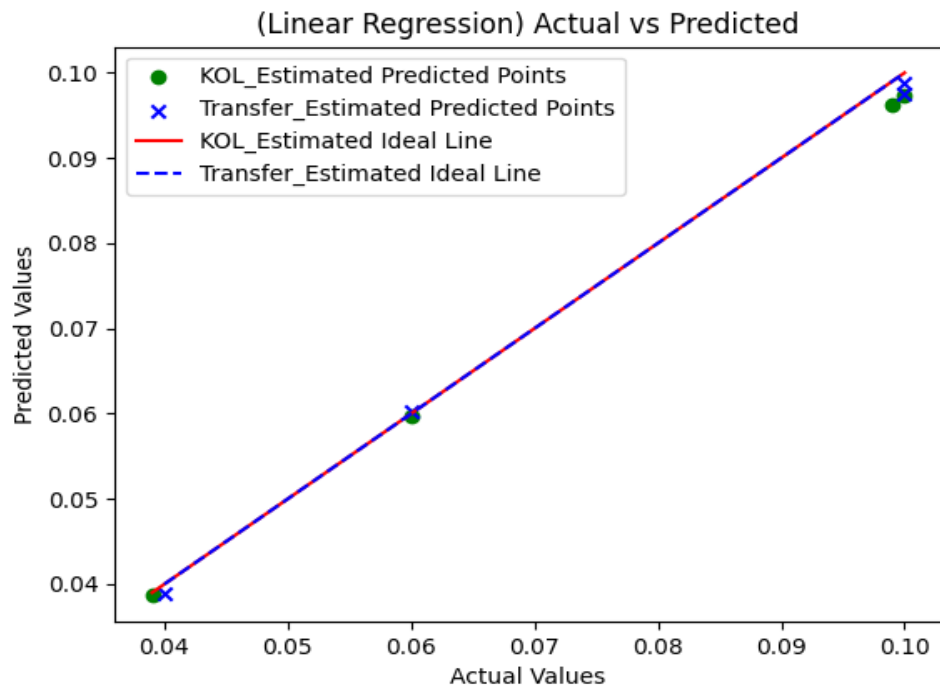
```
{'copy_X': True, 'fit_intercept': False, 'n_jobs': None, 'normalize':  
'deprecated', 'positive': False}
```

1.2 模型评分

原始模型评分：0.9945742997429028

调参后模型评分：0.9957302544584204

1.3 模型预测结果可视化



2. 决策树回归_KOL_Estimated模型

2.1 参数

原始模型参数：

```
{'ccp_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None,
'max_features': None, 'max_leaf_nodes': None, 'min_impurity_decrease': 0.0,
'min_samples_leaf': 1, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0,
'random_state': None, 'splitter': 'best'}
```

调参后模型参数：

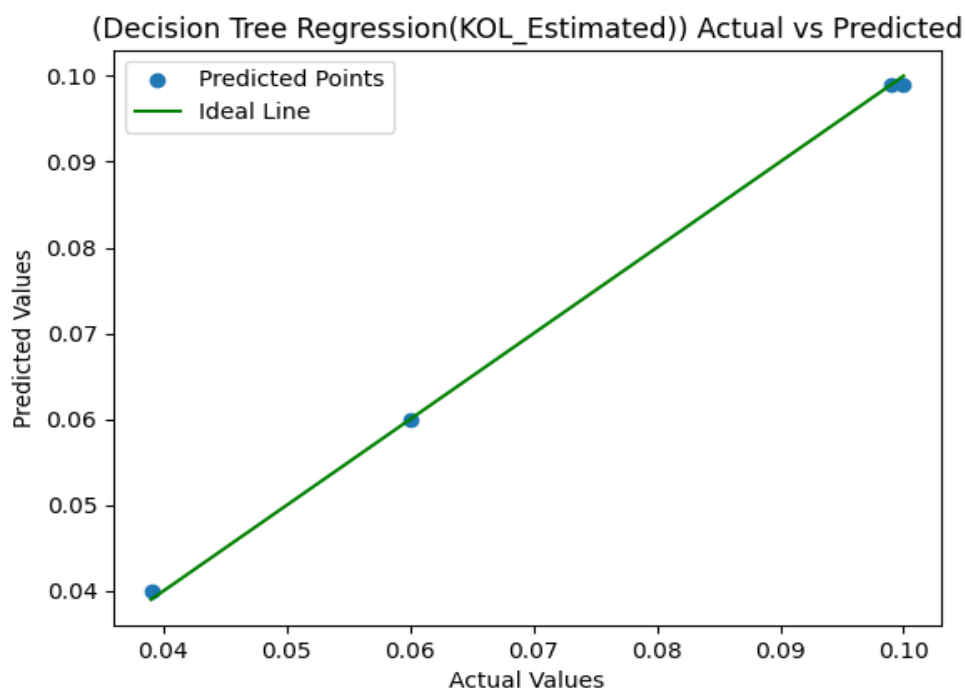
```
{'memory': None, 'steps': [('select', SelectKBest(k=3)), ('model',
DecisionTreeRegressor())], 'verbose': False, 'select': SelectKBest(k=3),
'model': DecisionTreeRegressor(), 'select__k': 3, 'select__score_func':
<function f_classif at 0x000001EB353AB5E0>, 'model__ccp_alpha': 0.0,
'model__criterion': 'squared_error', 'model__max_depth': None,
'model__max_features': None, 'model__max_leaf_nodes': None,
'model__min_impurity_decrease': 0.0, 'model__min_samples_leaf': 1,
'model__min_samples_split': 2, 'model__min_weight_fraction_leaf': 0.0,
'model__random_state': None, 'model__splitter': 'best'}
```

2.2 模型评分

原始模型评分：0.6905549430356486

调参后模型评分：0.9992649761117236

2.3 模型预测结果可视化



3. 决策树回归_Transfer_Estimated模型

3.1 参数

原始模型参数：

```
{'ccp_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None,
'max_features': None, 'max_leaf_nodes': None, 'min_impurity_decrease': 0.0,
'min_samples_leaf': 1, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0,
'random_state': None, 'splitter': 'best'}
```

调参后模型参数：

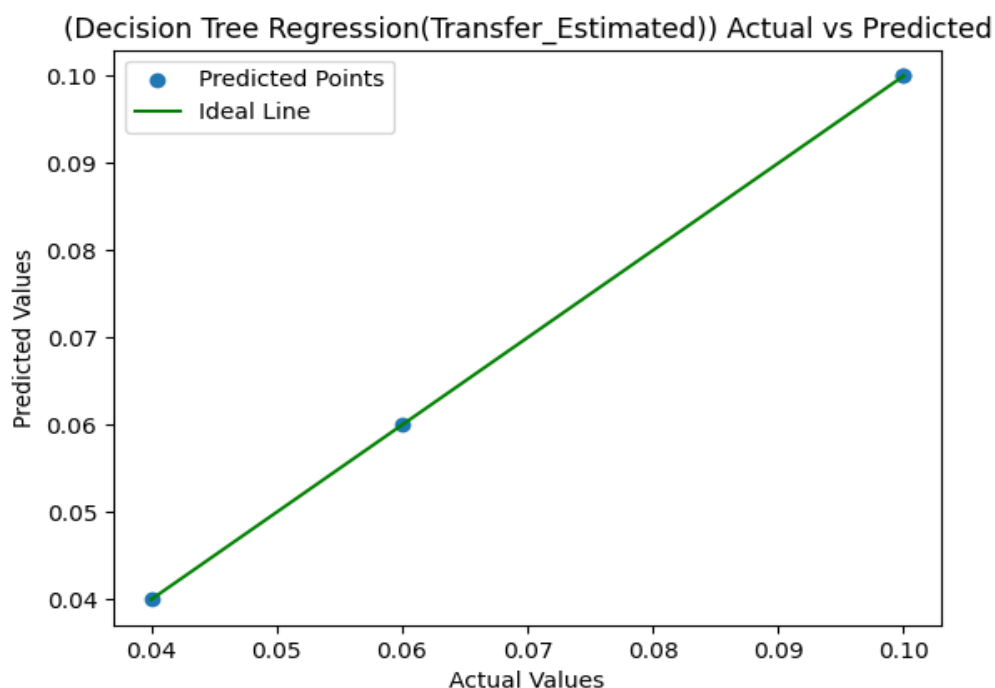
```
{'memory': None, 'steps': [('select', SelectKBest(k=3)), ('model',
DecisionTreeRegressor())], 'verbose': False, 'select': SelectKBest(k=3),
'model': DecisionTreeRegressor(), 'select__k': 3, 'select__score_func':
<function f_classif at 0x000001EB353AB5E0>, 'model__ccp_alpha': 0.0,
'model__criterion': 'squared_error', 'model__max_depth': None,
'model__max_features': None, 'model__max_leaf_nodes': None,
'model__min_impurity_decrease': 0.0, 'model__min_samples_leaf': 1,
'model__min_samples_split': 2, 'model__min_weight_fraction_leaf': 0.0,
'model__random_state': None, 'model__splitter': 'best'}
```

3.2 模型评分

原始模型评分：0.7037037037037036

调参后模型评分：1.0

3.3 模型预测结果可视化



4. 岭回归模型

4.1 参数

原始模型参数：

```
{'alpha': 1.0, 'copy_X': True, 'fit_intercept': True, 'max_iter': None, 'normalize': 'deprecated', 'positive': False, 'random_state': None, 'solver': 'auto', 'tol': 0.001}
```

调参后模型参数：

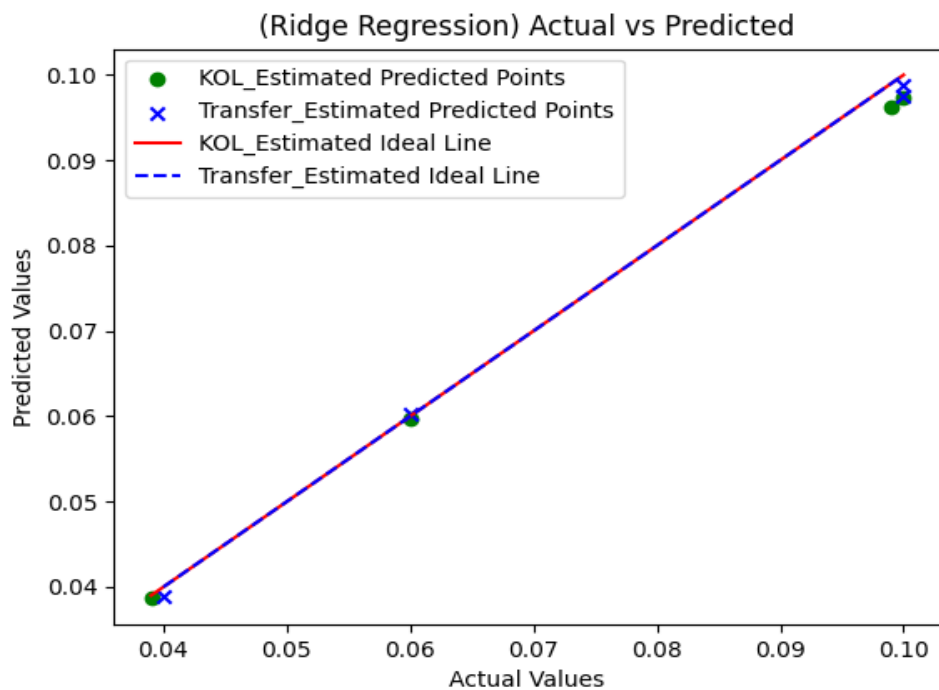
```
{'alpha': 1000, 'copy_X': True, 'fit_intercept': False, 'max_iter': None, 'normalize': 'deprecated', 'positive': False, 'random_state': None, 'solver': 'auto', 'tol': 0.001}
```

4.2 模型评分

原始模型评分：0.9945743118891728

调参后模型评分：0.9957395322800429

4.3 模型预测结果可视化



5. 拉索回归模型

5.1 参数

原始模型参数：

```
{'alpha': 1, 'copy_X': True, 'fit_intercept': True, 'max_iter': 12000,  
'normalize': 'deprecated', 'positive': False, 'precompute': False,  
'random_state': None, 'selection': 'cyclic', 'tol': 0.0001, 'warm_start': False}
```

调参后模型参数：

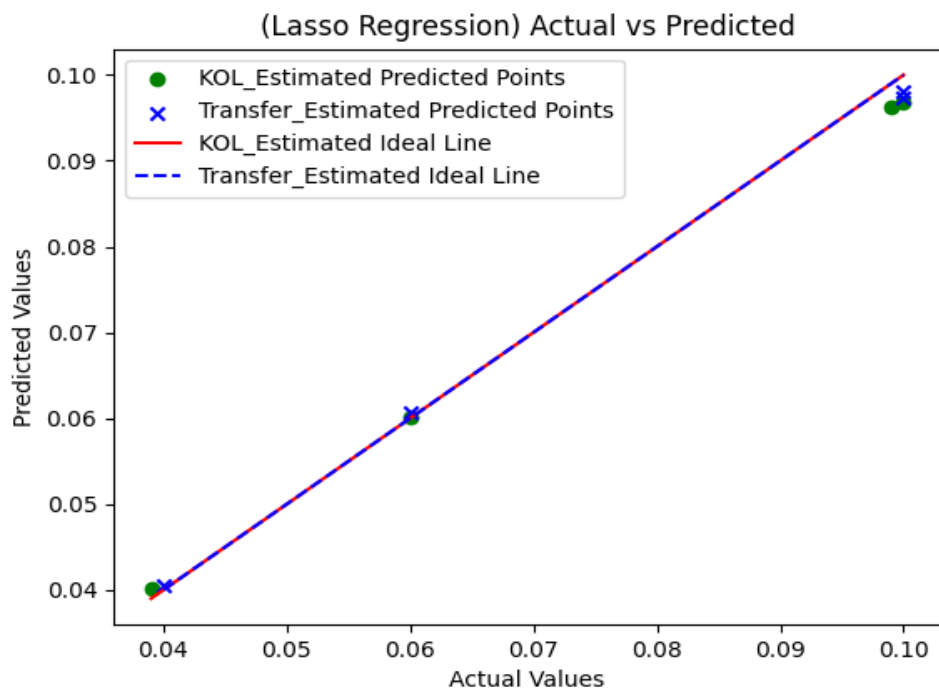
```
{'alpha': 1, 'copy_X': True, 'fit_intercept': True, 'max_iter': 12000,  
'normalize': 'deprecated', 'positive': False, 'precompute': False,  
'random_state': None, 'selection': 'cyclic', 'tol': 0.0005, 'warm_start': False}
```

5.2 模型评分

原始模型评分：0.9918964396498001

调参后模型评分：0.9943352461979191

5.3 模型预测结果可视化



6. SVM回归_KOL_Estimated模型

6.1 参数

原始模型参数：

```
{'C': 1.0, 'cache_size': 200, 'coef0': 0.0, 'degree': 3, 'epsilon': 0.1, 'gamma': 'scale', 'kernel': 'rbf', 'max_iter': -1, 'shrinking': True, 'tol': 0.001, 'verbose': False}
```

调参后模型参数：

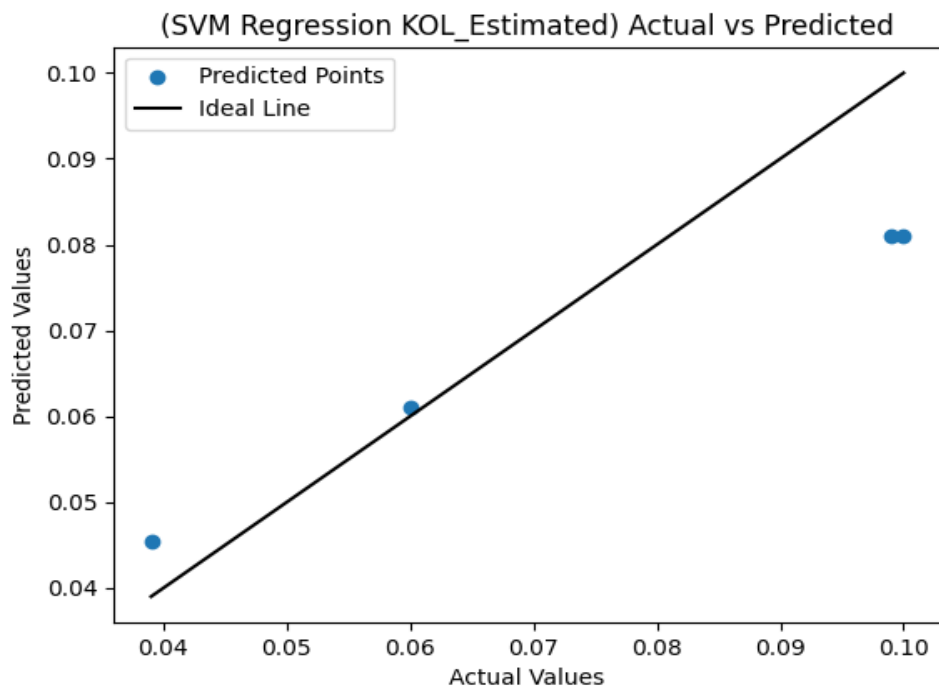
```
{'C': 100.0, 'cache_size': 150, 'coef0': 1.0, 'degree': 2, 'epsilon': 0.01, 'gamma': 'scale', 'kernel': 'sigmoid', 'max_iter': -1, 'shrinking': True, 'tol': 0.001, 'verbose': False}
```

6.2 模型评分

原始模型评分：-0.33076074972436587

调参后模型评分：0.7323426028532011

6.3 模型预测结果可视化



7. SVM回归_Transfer_Estimated模型

7.1 参数

原始模型参数：

```
{'C': 1.0, 'cache_size': 200, 'coef0': 0.0, 'degree': 3, 'epsilon': 0.1, 'gamma': 'scale', 'kernel': 'rbf', 'max_iter': -1, 'shrinking': True, 'tol': 0.001, 'verbose': False}
```

调参后模型参数：

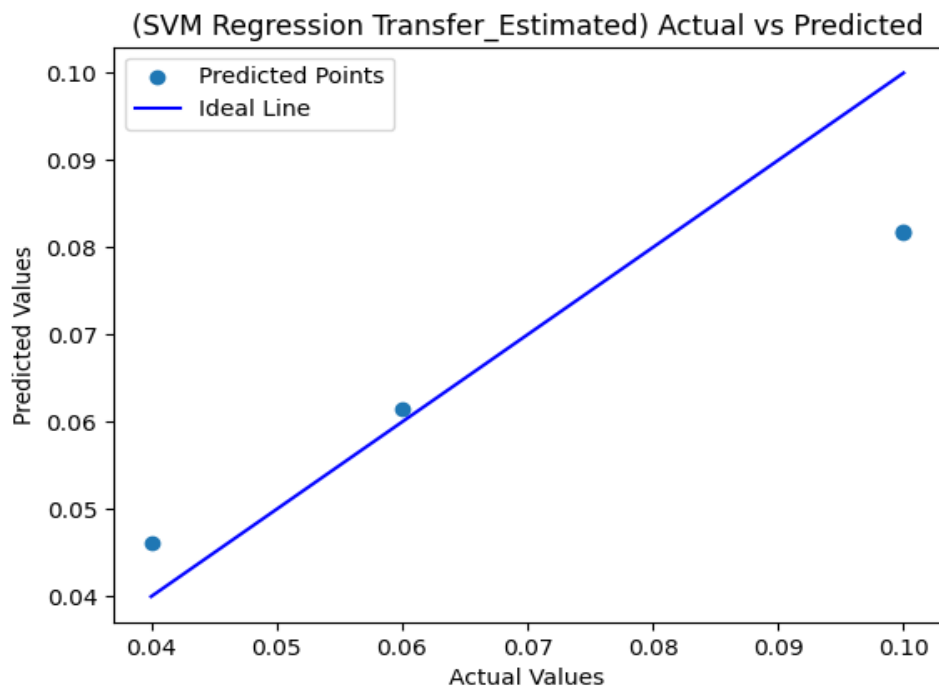
```
{'C': 100.0, 'cache_size': 150, 'coef0': 1.0, 'degree': 2, 'epsilon': 0.01, 'gamma': 'scale', 'kernel': 'sigmoid', 'max_iter': -1, 'shrinking': True, 'tol': 0.001, 'verbose': False}
```

7.2 模型评分

原始模型评分：-0.33333333333333304

调参后模型评分：0.7399541683745864

7.3 模型预测结果可视化



平均均方误差与均方误差汇总

	调参前平均均方 误差	调参后平均均方 误差	调参前均方误差 KOL_Estimated	调参后均方误差 KOL_Estimated	调参前均方误差 Transfer_Estimated	调参后均方误差 Transfer_Estimated
线性回归	7.52340369e-06	8.16179358e-06	4.23516259e-06	3.64170088e-06	3.12221869e-06	2.15056134e-06
决策树回归 _KOL_Estimated	4.74500000e-04	3.00000000e-07	3.10500000e-04	5.00000000e-07	nan	nan
决策树回归 _Transfer_Estimated	5.80000000e-04	0.00000000e+00	nan	nan	2.00000000e-04	0.00000000e+00
岭回归	7.52340432e-06	8.16171113e-06	4.23515493e-06	3.63587744e-06	3.12220989e-06	2.14381477e-06
拉索回归	4.15462703e-06	3.96182674e-06	6.68106003e-06	4.74810700e-06	4.31030920e-06	2.93595534e-06
SVM回归 _KOL_Estimated	8.00300000e-04	1.20653512e-04	9.05250000e-04	1.82073944e-04	nan	nan
SVM回归 _Transfer_Estimated	8.00000000e-04	1.21861295e-04	nan	nan	9.00000000e-04	1.75530936e-04

模型评分汇总

	训练前评分	训练后评分
线性回归	9.94574300e-01	9.95730254e-01
决策树回归_KOL_Estimated	6.90554943e-01	9.99264976e-01
决策树回归_Transfer_Estimated	7.03703704e-01	1.00000000e+00
岭回归	9.94574312e-01	9.95739532e-01
拉索回归	9.91896440e-01	9.94335246e-01
SVM回归_KOL_Estimated	-3.30760750e-01	7.32342603e-01
SVM回归_Transfer_Estimated	-3.33333333e-01	7.39954168e-01

模型选择结论：
决策树回归模型是最适合的模型。