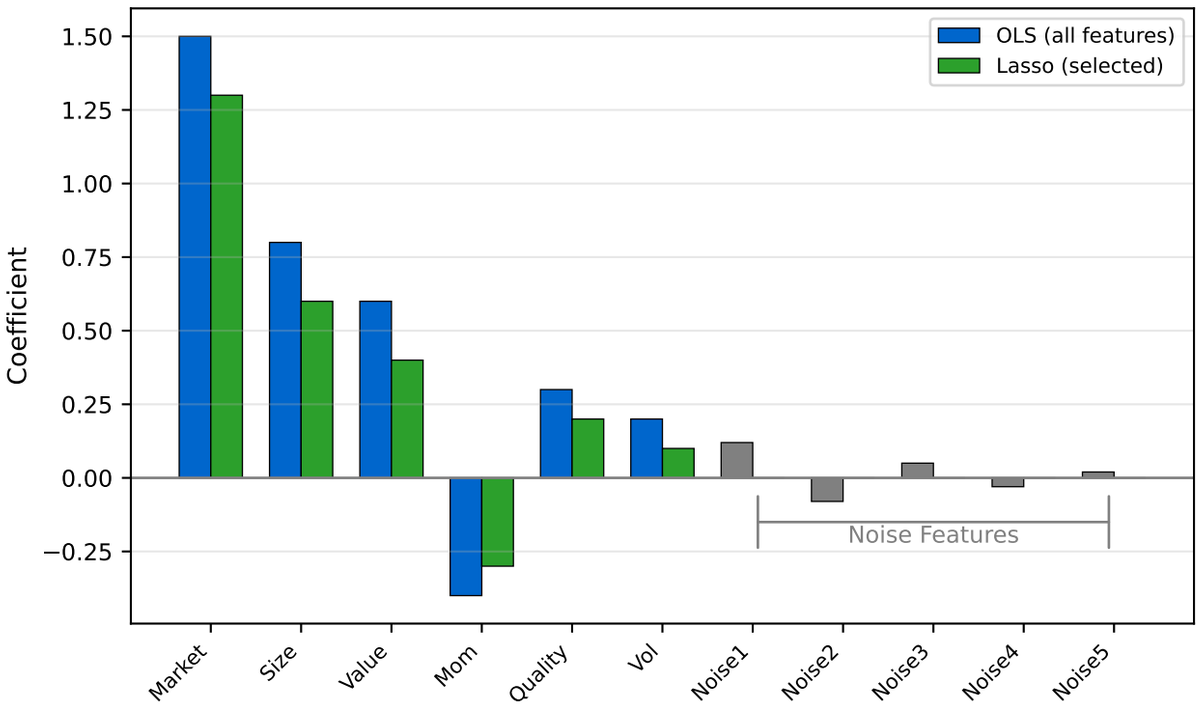
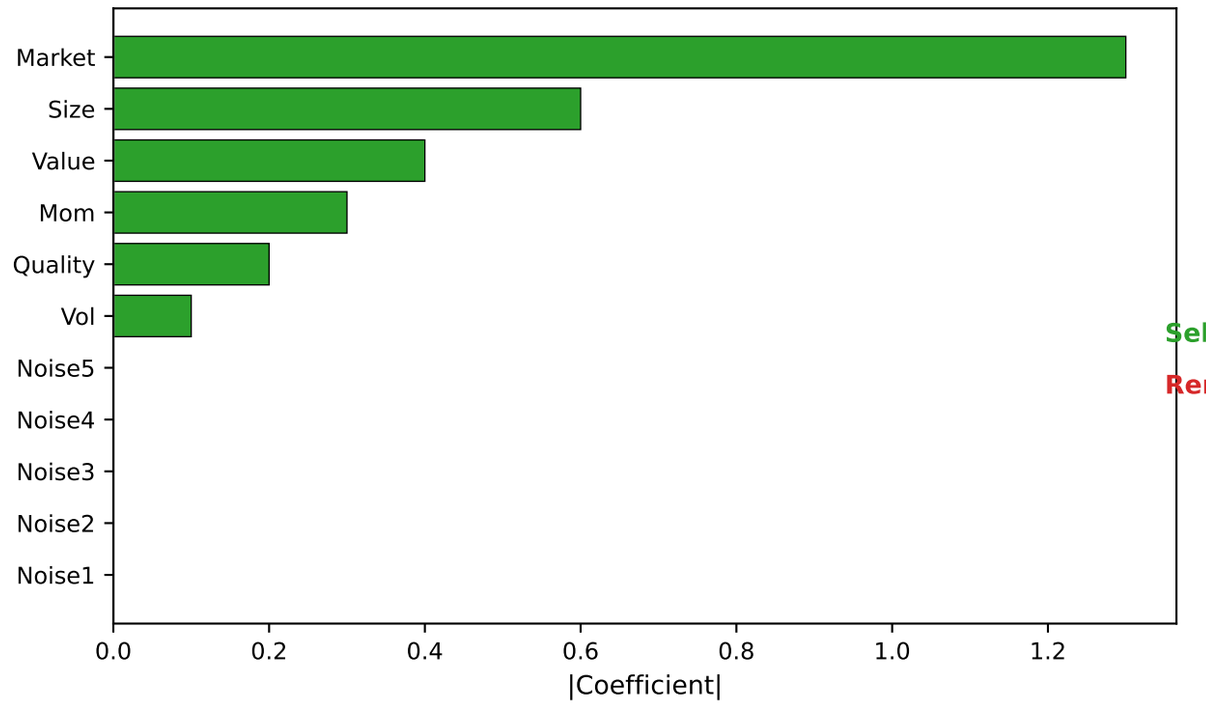
Feature Selection with Lasso

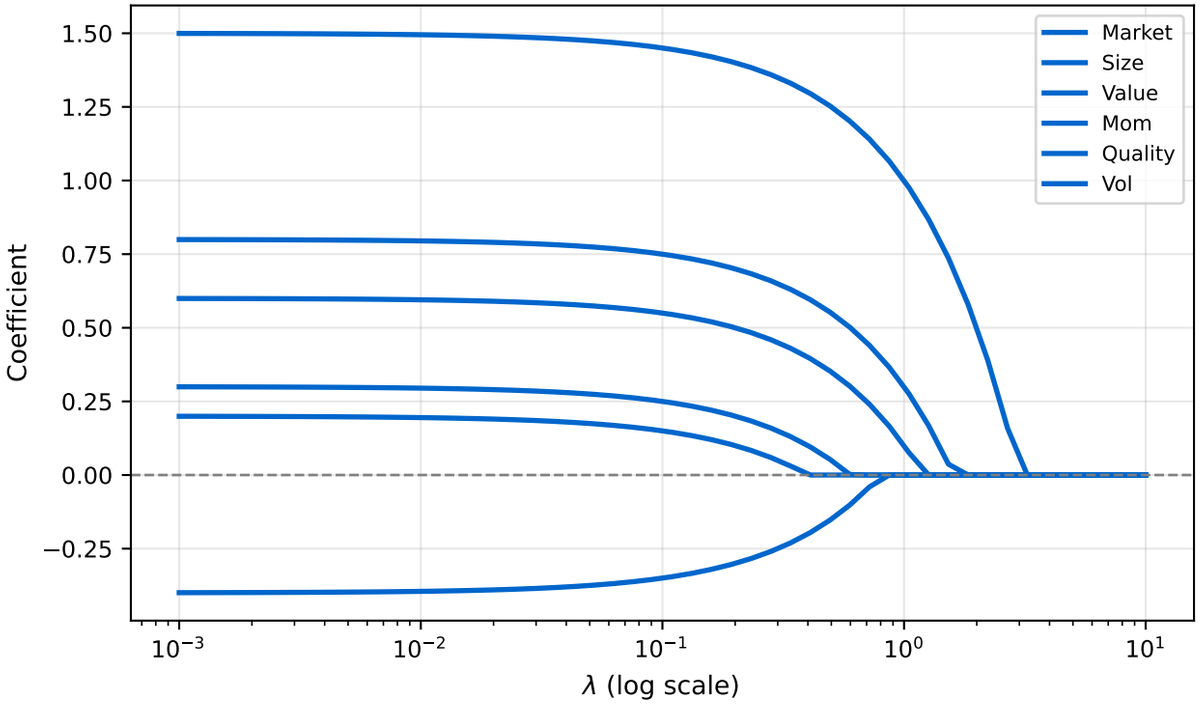
OLS Keeps Noise, Lasso Removes It



Feature Importance (|Lasso Coefficient|)



Lasso Path: Relevant Features Persist



Feature Selection Summary

```
LASSO FOR FEATURE SELECTION

Why use Lasso for feature selection?
- Automatically identifies relevant features
- Removes noise without manual filtering
- Produces interpretable sparse models

Process:
1. Fit Lasso with cross-validated lambda
2. Features with coef = 0 are removed
3. Features with coef != 0 are selected

Code:
from sklearn.linear_model import LassoCV

# Fit with automatic lambda selection
lasso = LassoCV(cv=5, random_state=42)
lasso.fit(X_train, y_train)

# Get selected features
selected = X.columns[lasso.coef_ != 0]
print(f"Selected features: {list(selected)}")

# Feature importance
importance = pd.Series(
    np.abs(lasso.coef_),
    index=X.columns
).sort_values(ascending=False)

Finance Application:
- Start with many potential factors
- Lasso selects truly predictive ones
- Build parsimonious factor model
```