Lesson 24: Factor Models

Data Science with Python – BSc Course

45 Minutes

Learning Objectives

The Problem: CAPM uses only market beta, but stocks also respond to size, value, and momentum. How do we capture multiple sources of systematic risk?

After this lesson, you will be able to:

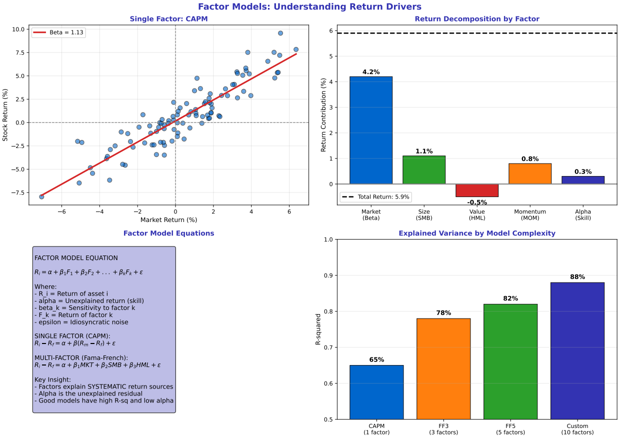
- Build multi-factor regression models
- Understand Fama-French factors (SMB, HML)
- Interpret factor loadings and alpha
- Create complete ML pipelines with sklearn

Finance Application: Decomposing returns into systematic factors for attribution

Factor Concept

From Single to Multiple Risk Sources

- CAPM: $R_i - R_f = \alpha + \beta_M(R_M - R_f)$ – only market risk
- Multi-factor: Add size, value, momentum as additional explanatory variables



Key insight: Different stocks have different exposures to different risk factors

The Classic Three-Factor Model

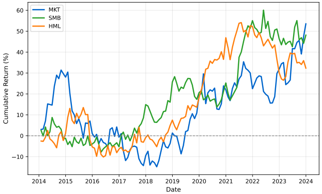
- **SMB** (Small Minus Big): Small caps outperform large caps historically
- **HML** (High Minus Low): Value stocks outperform growth stocks

Fama-French Factor Model

Factor Definitions

```
FAMA-FRENCH THREE-FACTOR MODEL (1993)

1. MKT (Market)
   - Market return minus risk-free rate
   - Rm - Rf
   - Captures broad market exposure

2. SMB (Small Minus Big)
   - Small cap return minus large cap return
   - Size effect: small stocks outperform
   - Based on market capitalization

3. HML (High Minus Low)
   - Value stocks minus growth stocks
   - Value effect: high B/M outperform
   - Based on Book-to-Market ratio

FAMA-FRENCH FIVE-FACTOR MODEL (2015)

4. RMW (Robust Minus Weak)
   - Profitable firms minus unprofitable
   - Profitability effect

5. CMA (Conservative Minus Aggressive)
   - Low investment minus high investment
   - Investment effect
```

Cumulative Factor Returns (10 Years)



Factor Statistics

FACTOR STATISTICS (Monthly)

Factor	Mean	Std	Sharpe (Ann.)
MKT	0.44%	4.15%	0.37
SMB	0.37%	3.02%	0.43
HML	0.28%	3.17%	0.31

CORRELATION MATRIX

	MKT	SMB	HML
MKT	1.00	0.10	-0.11
SMB	0.10	1.00	0.13
HML	-0.11	0.13	1.00

Key: Low correlations = good diversification

Python Implementation

Loading Fama-French Data with pandas

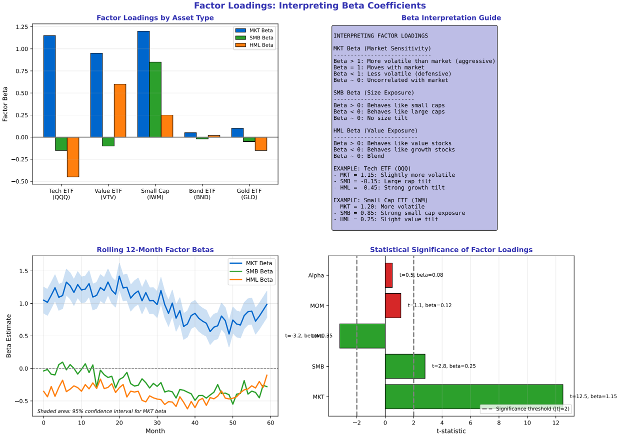
```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression

# Load Fama-French factors (from Ken French website)
# Format: Date, Mkt-RF, SMB, HML, RF (in %)
ff_data = pd.read_csv('F-F_Research_Data_Factors.csv',
                      skiprows=3)
ff_data['Date'] = pd.to_datetime(ff_data['Date'],
                                 format='%Y%m')

# Load stock returns
stock_returns = pd.read_csv('stock_returns.csv')

# Merge and prepare for regression
merged = pd.merge(stock_returns, ff_data, on='Date')
merged['excess_return'] = merged['return'] - merged['RF']

# Fit FF3 model
X = merged[['Mkt-RF', 'SMB', 'HML']]
y = merged['excess_return']

model = LinearRegression()
model.fit(X, y)

print(f"Alpha: {model.intercept_:.4f}")
print(f"MKT beta: {model.coef_[0]:.3f}")
print(f"SMB beta: {model.coef_[1]:.3f}")
print(f"HML beta: {model.coef_[2]:.3f}")
```

Nobel Prize (2013): Fama showed these factors explain returns better than CAPM alone

Factor Loadings

How Much Exposure Does a Stock Have?

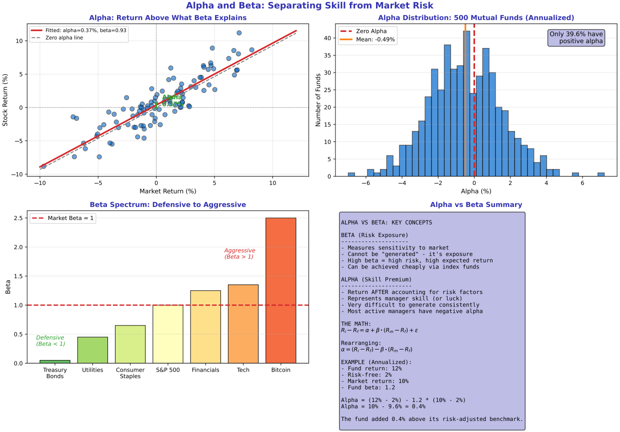
- Each stock has different sensitivities to each factor
- Loading = regression coefficient on that factor



Example: TSLA has high market beta but negative HML (growth stock, not value)

Separating Skill from Risk Exposure

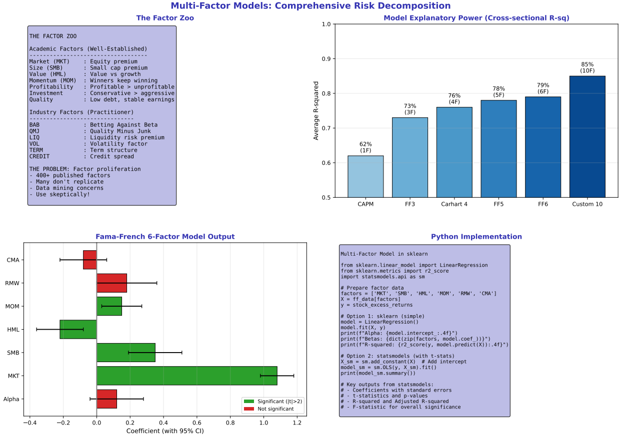
- α: Return not explained by factors (manager skill or mispricing)
- Multi-factor alpha is "purer" than CAPM alpha



Industry standard: Report alpha after controlling for Fama-French factors

Implementation in Python

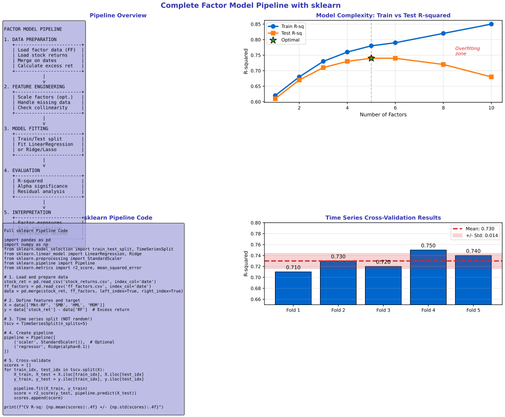
- Same sklearn API: `LinearRegression().fit(X, y)` where X has multiple columns
- Each coefficient is a factor loading



X matrix columns: [Mkt-RF, SMB, HML] – intercept is alpha

Combining Preprocessing and Modeling

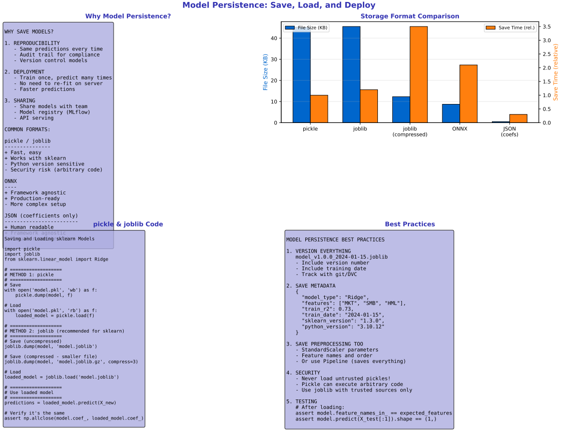
- `Pipeline([('scaler', StandardScaler()), ('reg', Ridge())])`
- Prevents data leakage: scaler fits only on training data



Best practice: Always use pipelines for reproducible workflows

Saving and Loading Trained Models

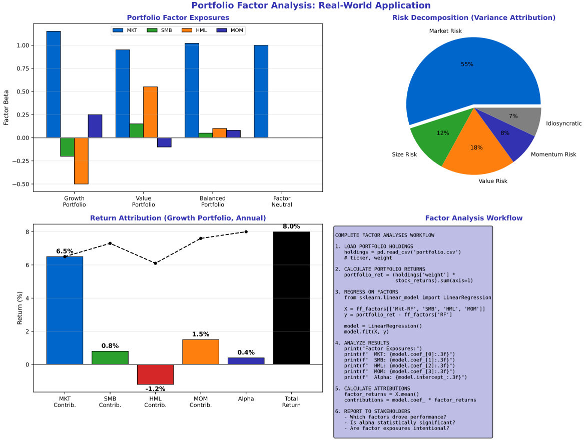
- `joblib.dump(model, 'model.pkl')` – save to disk
- `model = joblib.load('model.pkl')` – reload for production



Deployment: Train once, deploy saved model to production

Where Did Your Returns Come From?

- Decompose portfolio returns into factor contributions
- Shows whether performance came from market timing, factor bets, or alpha



Risk management: Understand your factor exposures before they surprise you

Hands-On Exercise (25 min)

Task: Build a Fama-French Factor Model

1. Download Fama-French 3-factor data from Kenneth French's website
2. Merge with your stock returns (AAPL or similar)
3. Fit multi-factor regression: stock returns vs [Mkt-RF, SMB, HML]
4. Interpret: What are the factor loadings? Is there alpha?
5. Compare R^2 to single-factor CAPM model

Deliverable: Table of factor loadings + comparison of R^2 values.

Extension: Add momentum (UMD) as a fourth factor – does R^2 improve?

Lesson Summary

Problem Solved: We can now decompose stock returns into multiple systematic factors, giving better risk attribution than CAPM alone.

Key Takeaways:

- Fama-French: Market + SMB (size) + HML (value)
- Factor loadings = regression coefficients on each factor
- Alpha after factors = true outperformance
- sklearn pipelines ensure reproducible workflows

Next Lesson: Classification (L25) – predicting categories instead of numbers

Memory: SMB = Small Minus Big (size), HML = High Minus Low (value)