DATA 201: Time Series Analysis
Linear Regression Model
Lecture 8: Size portfolios

Lulu Wang

Data Analytics
Dickinson College

2/14/2025

Introduction

CAPM and Fama-French

What is Market Capitalization

Size portfolios

Objectives

- Review the Linear Regression Model (LRM)
- Discuss its estimation in Python
- Explore applications in financial data

I will review the regression model in broad terms and more details can be found in an introductory statistics/econometrics textbook, such as:

- Stock and Watson, *Introduction to Econometrics*, Pearson
- Wooldridge, *Introductory Econometrics*, A Modern Approach, South-Western

The Linear Regression Model (LRM)

- The linear regression model is given by:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t,$$

where:

  - Y_t: dependent variable at time t
  - X_t: independent variable / factor / predictor at time t
  - β_0, β_1: coefficients to be estimated
  - ε_t: error term (mean zero, variance σ^2)

- The **expected** (or average) value of Y_t given X_t is:

$$E(Y_t \mid X_t) = \beta_0 + \beta_1 X_t.$$

CAPM

- **Interpretation**:
  - β_0: the expected value of Y_t when $X_t = 0$
  - β_1: the expected change of Y_t for a unit change of X_t
- The Capital Asset Pricing Model (CAPM) is an example of an LRM:

$$R_t^i \;=\; \beta_0 \;+\; \beta_1 \, R_t^{\mathrm{MKT}} + \varepsilon_t,$$

where R_t^i and R_t^{MKT} represent the excess stock and market returns, respectively.

LRM with Multiple Independent Variables

- Typically, we have several independent variables (factors) relevant to explain the dependent variable.
- We extend the Linear Regression Model (LRM) to include independent variables denoted as $X_{k,t}$ for $k = 1, \ldots, K$:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_K X_{K,t} + \epsilon_t$$

- The model is estimated by OLS, making the formula more complex than the single regressor case.

Fama-French Three-Factor Model

- In asset pricing, independent variables are typically referred to as risk factors.
- Fama and French (1993) extend CAPM with two additional factors:
    - **SMB** (Small-minus-Big): Return spread between small and large capitalization stocks.
    - **HML** (High-minus-Low): Return spread between high and low Book-to-Market ratio stocks (value vs. growth stocks).

FF3 Model Equation

- The Fama-French Three-Factor model is defined as:

$$R_t^i = \beta_0 + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \epsilon_t$$

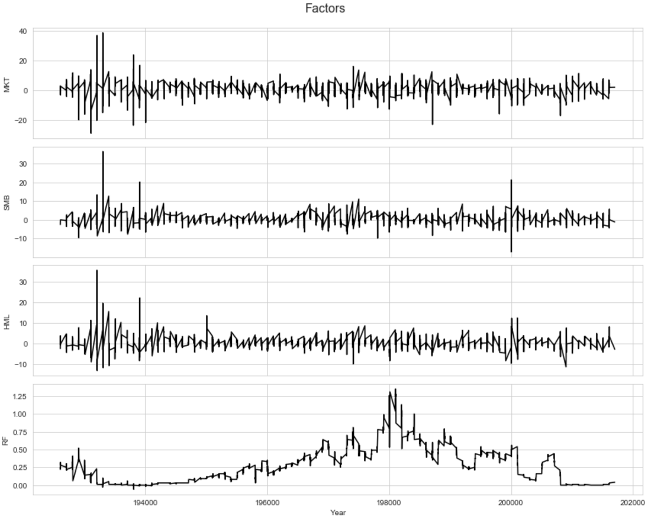
- The goal is to decompose portfolio return R_t^i into exposures to:
  - **Systematic risk**: $\beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t$
  - **Idiosyncratic risk**: ϵ_t
  - **Risk-adjusted return**: β_0

Economic Interpretation of FF3 Factors

- **Market Risk (MKT_t)**: Captures overall market movement.
- **SMB (Size Factor)**:
  - Small firms tend to have higher expected returns than large firms.
  - Related to firm size premium in asset pricing.
- **HML (Value Factor)**:
  - Value stocks (high Book-to-Market) outperform growth stocks (low Book-to-Market).
  - Reflects differences in risk exposure between value and growth stocks.

A first look at FF factors

- The data for the factors are downloadable at Ken French webpage (1926-2017)



A first look at FF factors(Cont.)

- Compute summary statistics:

	Mean	Std Dev	Skewness	Kurtosis	Max	Min
MKT	0.65	5.37	0.19	7.79	38.85	-29.13
SMB	0.21	3.20	1.89	19.14	36.56	-17.20
HML	0.39	3.52	2.31	20.19	35.61	-13.11
RF	0.28	0.25	1.06	1.25	1.35	-0.06

- Correlation matrix among the Fama-French factors (including the riskfree rate):

	MKT	SMB	HML	RF
MKT	1.00	0.32	0.25	-0.07
SMB	0.32	1.00	0.12	-0.05
HML	0.25	0.12	1.00	0.02
RF	-0.07	-0.05	0.02	1.00

What is Market Capitalization?

- **Definition:**
  - Market Capitalization **(Market Cap) = Stock Price $\times$ Number of Outstanding Shares**
  - Represents the **total value of a company's equity** in the stock market.
- **Example:**
  - Tesla (TSLA): Stock Price = \$350, Shares Outstanding = 3.2B
    - Market Cap = \$1.1T (Large Cap)
  - Rivian (RIVN): Stock Price = \$15, Shares Outstanding = 900M
    - Market Cap = \$13.5B (Mid Cap)

Market Capitalization Categories

Companies are classified into different **market cap tiers**:

Category	Market Cap Range	Example Companies
Large-Cap	\$10B+	Apple(AAPL), Microsoft(MSFT)
Mid-Cap	\$2B – \$10B	Zoom(ZM), Etsy(ETSY)
Small-Cap	\$300M – \$2B	Upstart(UPST), Beyond Meat(BYN
Micro-Cap	Below \$300M	Nano-cap biotech stocks

Key Observation:

- **Large caps** are well-established and less volatile.
- **Small caps** can grow faster but carry more risk.

Market Cap and Stock Performance

- **Historical Performance:**
  - Small caps tend to outperform large caps over long periods.
  - Large caps are more stable and resilient during downturns.
- **Example:Russell 2000 vs. S&P 500 (2000-2023)**
  - **The Russell 2000 (small-cap index)** has averaged **9-10%** annual returns, outperforming large caps in growth phases.
  - **The S&P 500 (large-cap index)** has averaged **7-8%** annual returns, with lower volatility.

Ken French Portfolio Dataset

- The webpage of Ken French provides historical returns of portfolios of stocks sorted by several indicators:
  - Market capitalization
  - Book-to-market ratio
  - Past 6-month returns
- The portfolio formation works as follows:
  1. Every June, all stocks listed in the US equity markets (**NYSE, NASDAQ**) are sorted based on the indicator (e.g., market capitalization).
  2. Portfolios are created that include the τ% of stocks with the lowest value of the indicator, τ% of stocks with the highest value, and intervals of τ% between them.
  3. The portfolio is held for one year and revised the following June.

Portfolio Categories and Sample Period

- The dataset provided by French includes three portfolio categories:
  - Lowest 30%, mid 40%, highest 30%
  - Portfolio of 20% of the stock universe (quintiles)
  - Portfolio of 10% of the stock universe (deciles)
- **Sample period:** The portfolios cover the period from **July 1926 to November 2017**.

Size portfolios

- Size portfolios are formed by sorting stocks on their market capitalization (= outstanding shares times stock price)
- The `size10` object contains the decile portfolios (10% of the stock universe in each portfolio)

	AV RET	STD DEV
Lo 10	1.116	9.913
Dec 2	0.990	8.738
Dec 3	0.983	7.938
Dec 4	0.946	7.455
Dec 5	0.892	7.034
Dec 6	0.908	6.796
Dec 7	0.834	6.404
Dec 8	0.806	6.128
Dec 9	0.736	5.791
Hi 10	0.613	5.042

- The portfolio of the 10% smallest cap stocks (`Lo 10`) earns an average monthly return of 1.116% while the portfolio of the 10% largest cap stocks (`Hi 10`) earns 0.613%
- What explains an over-performance by 0.503%? Could it be that the small cap portfolios earn higher returns because they are more risky?

CAPM and size portfolios

- In the context of the CAPM higher risk means larger exposure/beta to the MKT factor.
- Let's estimate the CAPM model and see if this hypothesis is confirmed in the data.

	const	MKT
Lo 10	0.177	1.418
Dec 2	0.070	1.388
Dec 3	0.105	1.326
Dec 4	0.109	1.263
Dec 5	0.078	1.227
Dec 6	0.108	1.207
Dec 7	0.073	1.149
Dec 8	0.068	1.114
Dec 9	0.031	1.064
Hi 10	-0.003	0.931

CAPM and size portfolios (Cont.)

- The MKT beta (slope) is equal to 1.407 for `Lo 10` and declines for larger cap portfolios toward 0.932 for the `Hi 10` portfolio.
- Hence, we do find that the small cap portfolio are more risky as measured by the MKT beta.
- Let's convert the table of estimates into a graph.

CAPM and size portfolios (Cont.)

- Are the estimated alpha (intercept) and beta (MKT slope) related?
- If the CAPM is the correct pricing model then the alphas should be equal to zero
- Larger alphas are observed for the small cap portfolios (low deciles)

CAPM and size portfolios (Cont.)

- Let's evaluate statistically the hypothesis that the alpha coefficients are equal to zero, that is, $H_0 : \alpha = 0$ against $H_1 : \alpha \neq 0$
- The t-statistics indicate that none of the alphas are significant at 5% (critical value 1.96), except for the 6th portfolio at 10% (critical value 1.645)level

Lo 10	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Hi 10
0.909	0.501	0.961	1.127	1.011	1.659	1.316	1.535	0.911	-0.139

- Although the alphas are mostly positive, statistically they are not different from zero
- What about the goodness-of-fit of these 10 regression models?

Lo 10	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Hi 10
0.584	0.721	0.797	0.820	0.869	0.901	0.920	0.944	0.964	0.973

Size portfolios: 3-factor model

- The CAPM seems to explain relatively well the size portfolios:
    - Beta are larger for the small cap portfolios that provide higher expected returns.
    - Alphas are positive and large for the small cap portfolios, although not statistically significant at 5%.
- What can we learn from estimating the 3-factor model?

$$R_t^i = \alpha + \beta_{MKT} R_t^{MKT} + \beta_{SMB} R_t^{SMB} + \beta_{HML} R_t^{HML} + \epsilon_t$$

- Are **SMB** and **HML** going to provide *explanatory power*?
- Is R^2 increasing by adding these two variables?
- Are the intercept and the slope coefficients changing by adding these two variables?

Size portfolios: 3-factor model (Cont.)

- Findings:
  - intercept estimates are mostly negative
  - MKT slopes are very close to 1 (compare with CAPM estimates!)
  - SMB slopes are large and positive for Lo 10 and decline toward zero for the largest cap portfolios
  - HML slopes are positive and declining moving from Lo 10 to Hi 10

	const	MKT	SMB	HML
Lo 10	-0.163	0.999	1.548	0.786
Dec 2	-0.166	1.068	1.275	0.490
Dec 3	-0.079	1.075	1.008	0.376
Dec 4	-0.050	1.045	0.881	0.320
Dec 5	-0.034	1.059	0.724	0.197
Dec 6	0.006	1.077	0.494	0.228
Dec 7	0.003	1.050	0.408	0.136
Dec 8	0.016	1.051	0.229	0.121
Dec 9	-0.005	1.033	0.065	0.112
Hi 10	0.023	0.977	-0.215	-0.034

Size portfolios: 3-factor model (Cont.)

- Comparing the adjusted R^2 of the CAPM and 3-factor models there is a remarkable increase in goodness-of-fit for the small-cap portfolios and a marginal improvements for the high deciles.
- Clearly the 3-factor model outperforms the CAPM in pricing these portfolios.

	Lo 10	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Hi 10
CAPM	0.584	0.721	0.797	0.820	0.869	0.901	0.920	0.944	0.964	0.973
3-FACTOR	0.893	0.961	0.977	0.975	0.979	0.965	0.964	0.962	0.970	0.991