Slide Deck #1:

Course Introduction

Analytics with Big Finance Data

- Stock trading records, financial statements, and other mandatory disclosures contain large amounts of financial data that provide the foundation for a variety of forward-looking decisions.

- Recent trends in Big Data and predictive analytics are revolutionizing the way analysts extract meaningful insights from these disclosures and applies financial theory to asset management.

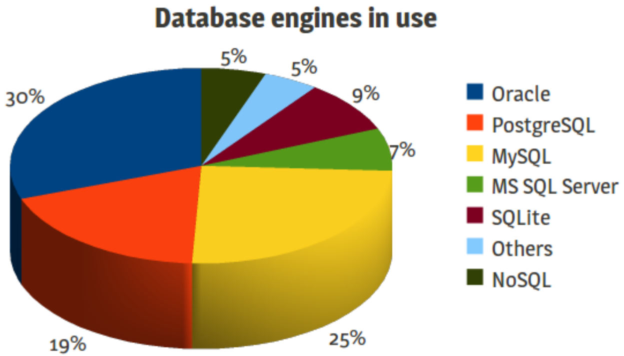
- This course teaches students the hands-on skills necessary to manipulate large-scale financial databases and build predictive models useful for strategic and investment decisions.

Why is SQL

- SQL is an essential in data science, and at the heart of all relational databases including Oracle, MySQL, PostgreSQL, etc.
- SQL can be invoked in many popular statistical packages including SAS, Python, R, and MATLAB--making it a "must know" for anyone interested in analyzing large datasets
- According to a Forbes July 2017 report, SQL is the second-most in-demand skill in data analysis (after "critical thinking")
  - The average SQL Developer salary, according to Indeed.com, was $92,000

Why is SQL

- Data is a core part of many financial and accounting analysis
  - Such as financial statements, stock trading records, Facebook profile, etc.
- SQL enables human to manage and program with Big Data
  - Other programming language, such as Python, can also handle data in smaller scale, but SQL is specialize in Big Data



Calculate ROA remotely

```
/* PC-SAS/Connect Communication Block  */
%let wrds = wrds.wharton.upenn.edu 4016;
options comamid=TCP remote=WRDS;
signon username=_prompt_;
/* Submit SAS code to WRDS Unix Server */
rsubmit;
    /* SAS CODE submitted to WRDS Sever    */
    proc sql;
    create table work.demo as

    select gvkey, tic, fyear, ni/at as roa

    from comp.funda

    where not missing(at) and consol="C" and indfmt="INDL" and
    datafmt="STD" and popsrc="D" and fyear=2019

    order by gvkey, fyear;
    quit;
/* Remote SAS CODE submission Ends */
end submit;
/* Sign off from PC-SAS/Connect    */
Signoff;
```

Some non-SAS coding

- PostgreSQL Data Extraction

```
comp = conn.raw_sql("""
                    select gvkey, tic, fyear, ni, at,
                      case when at=0 then null else ni/at end as roa
                    from comp.funda
                    where at is not null and indfmt='INDL' and datafmt='STD'
                      and popsrc='D' and consol='C' and fyear=2019
                    order by gvkey, fyear
                    """)
```

- Python (Pandas) Data Extraction

```
msi=pd.read_sas("/wrds/crsp/sasdata/a_stock/msi.sas7bdat")
```

CRSP Stocks

- The CRSP U.S. Stock database contains end-of-day and month-end prices on primary listings in major stock exchanges, along with basic market indices.
- CRSP databases are characterized by their comprehensive corporate action information and highly accurate total return calculations.
- CRSP U.S. Stock database **contains the following information**:
  - Price and quote data (e.g. Open, close, bid/low, ask/high, Trade-Volume).
  - Holding period returns with and without dividends.
  - Excess returns and other derived data items.
  - Shares outstanding.

Compustat Fundamentals

- Compustat provides more than **500 company-level fundamentals**, including items such as Income Statements, Balance Sheets, and Flow of Funds.
- It also offers an even larger number of supplemental data items for more than **47,000 active and 37,000 inactive companies**.
- Compustat primarily draws its US **data from SEC filings**, which it standardizes to allow for better comparisons
- Compustat is traditionally split between North American and Global (rest of world)

I/B/E/S Estimates

- I/B/E/S, the Institutional Brokers' Estimate System, is a database from Thomson Reuters.
- It is an historical earnings estimate database containing analyst estimates for more than 20 forecast measures - including EPS (earnings per share), price targets - available on both consensus and detailed levels, covering both U.S. and international companies.
- The database also includes buy-hold-sell recommendations.
- I/B/E/S History is a comprehensive expectation database and is ideal for research and back-testing of investment theories and proprietary models.

Slide Deck #1 Supplementary:

Things you need to know

Stock and Firm Identifiers

- To deal with those problems, many commercial databases create permanent identifiers. However, there are a caveat you need to pay attention:
  - Stock Identifier (Many companies have more than one classes of shares)
    - PERMNO for CRSP and {GVKEY + IID} of Compustat
  - Company Identifier
    - PERMCO for CRSP and GVKEY for Compustat

Date	CUSIP	Ticker	Name	CRSP Security Identifier (PERMNO)	S&P Identifier (GVKEY)	S&P Security-level Sub-Identifier (IID)
20060630	67461T10	RHEO	Occulogix Inc	90488	161071	01
20060901	67461T10	OCCX	Occulogix Inc	90488	161071	01
20081106	67461T20	OCCX	Occulogix Inc	90488	161071	01
20090116	67461T20	TEAR	Occulogix Inc	90488	161071	01
20100521	87819310	TEAR	Tearlab Corp	90488	161071	01

Data Frequency and Structures

- Stock Price Data are from real trading systems:
  - Tick-by-tick (HFT), daily (Event Study), Monthly and Annual(Back Testers)
- Fundamental (Accounting) Data are from Regulatory Bodies, such as the SEC and NYSE
  - Quarterly (partitioner) and Annually (academic)

	CRSP	**Compustat North American**
Time	From 1925	From 1950
Universe	Firms listed in USA Major Exchanges • NYSE, NASDAQ, AMEX, and ARCA	USA (10K filers) and Canada Firms •listed in major USA or Canada Exchanges, •listed in regional exchanges, •with shares traded as OTC, or •with certain amount of public bonds
Frequency	Daily or Monthly	Quarterly and Annually
Inclusion	Securities are included since IPO	Traditional, Compustat includes a firm only after a few years the firm publicly disclose its financial statement.

Return Calculation

$$\textbf{Return} = \frac{\textbf{Ending Price} \; \overset{\text{+ Dividends}}{} \; - \; \textbf{Beginning Price}}{\textbf{Beginning Price}}$$

- CRSP Return:

$$Ret_t = \frac{P_t + Div_{t-1,t} - P_{t-1}}{P_{t-1}}$$

Road Map

Highlights of WRDS Structure

- Web Interface
  - CRSP (Daily and Monthly)
    - Company Look Up
    - Trading Record Pull Down
  - Compustat (Annual)
    - Company Look Up
    - Annual Data Pull Down

Alternative Data Venue

- WRDS SAS Studio
  - File Structure
  - Data Download and Upload
- Python on WRDS
  - WRDS Package
  - Two ways to obtain data

Slide Deck #2:

Asset Management

Efficient Market Hypothesis

- Fama (1970) : "A market in which prices always *fully reflect* available information is called *efficient*."
- In layman's terms:
  - The pricing mechanism of markets —or trading — efficiently reflected all known information about stocks not clear
  - If there were any significant details known about a company, someone would discover it eventually and act on that information.
- Implication:
  - stock-picking was a futile exercise (by Fama)
    无效的

Near-Efficient Market

- Grossman and Stiglitz (1980) have shown abnormal returns will exist if there are costs of gathering and processing information.
- In layman's terms:
  - ***Suppose it is costly to collect and trade on information (as in the real world)***
  - **WHY** would anyone ever invest in gathering the information, if all information is in the price already?
  - **HOW** can information be reflected in security prices so that markets are efficient, if no one invests in gathering the information?

A Foundation of Finance Research

- Despite the modern notion that markets are not perfectly efficient, tests of the EMH are still important because they allow investors to gauge where they may make excess returns.

  判定

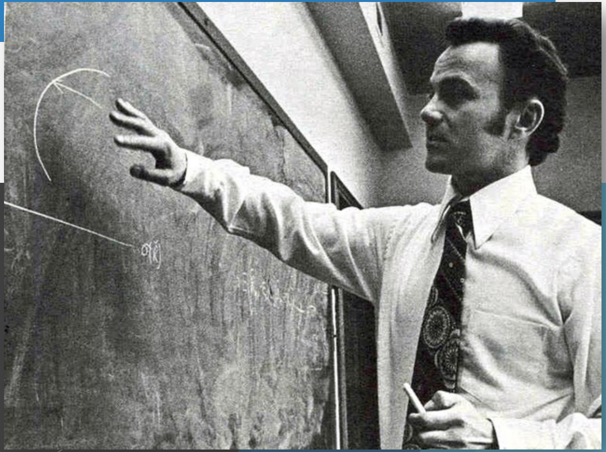
- The efficient market hypothesis has been recognized as one of the basic building blocks of modern financial economics

- EMH provides the foundation of thoughts for this course

相互作用 穿透

"Finance is the most successful branch of economics in terms of theory and empirical work, the interplay between the two, and the penetration of financial research into other areas of economics and real-world applications."

"I have been doing research in finance almost since its start, when Markowitz (1952, 1959) and Modigliani & Miller (1958) set the field on the path to becoming a serious scientific discipline."



My Life in Finance

Eugene F. Fama

Investment Motives

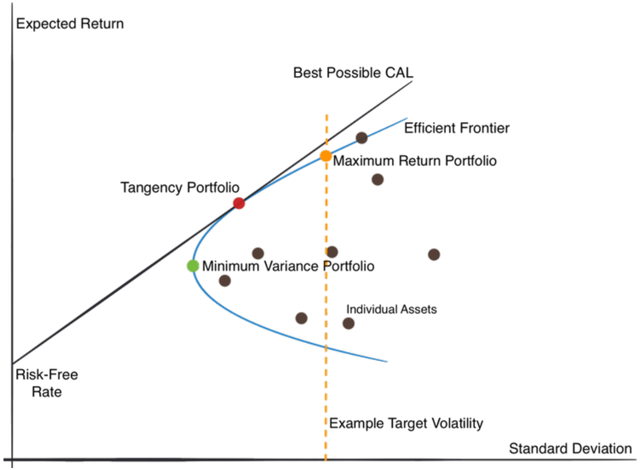
- Apply to all investment assets such as
  - Corporate/Government Bond
  - Crypto Currency
  - ETFs and Mutual Funds
  - Real Estates

Portfolio can help reduce risk

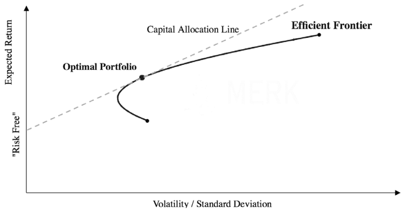
- Things that affect stock prices
  - Industry growth, firm profitability, regulatory environment, etc.
  - The greater the dissimilarity between the two firms' operating environments, the lower the correlation between their stock performances.
  - The greater the diversification of a portfolio, the lower the total variability (variance) the portfolio can accomplish.
    - Negative correlation between stock returns leads to a variance reduction
- Systematic and diversifiable risk
  - Systematic risk such as economic conditions, and natural disasters that affect all the firms in similar way
  - Diversifiable risk: individual firm specific (idiosyncratic) risk

Efficient Frontiers and Market

- When we assume investors are rational, risk averse, and prefer higher returns, we have a optimization problem find a set of portfolio weight vector to minimize the risk (standard deviation) and maximize returns.



Foundation of Modern Finance



$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$$

Harry M. Markowitz

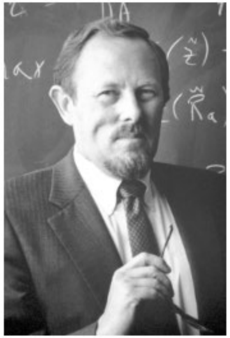
The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1990

Prize motivation: "for their pioneering work in the theory of financial economics"
Born: 24 August 1927, Chicago, IL, USA

William F. Sharpe

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1990

Prize motivation: "for their pioneering work in the theory of financial economics"
Born: 16 June 1934, Boston, MA, USA



Factor Theory

$r_i - r_f$

- Assets earn **risk premiums** (return in excess of the risk-free rate) because they are exposed to underlying **factor risks.**
- CAPM (first theory of factor risk) states that assets that crash when the market loses money are risky, and therefore must reward their holders with high **risk premiums.**
  - Safe(risky) stocks lose less (more) than the market portfolio in a falling market and gain less (more) in a bullish market.
  - CAPM is developed from Markowitz's theory with a same set of assumptions
- Multifactor models capture multiple definitions of bad times across many factors and states of nature.

A Factor Interpretation of CAPM

- CAPM states that all assets have different exposures to the market factor and the greater the exposure, the higher the risk premium.

$$\underbrace{\boldsymbol{R_i - r_f}}_{\text{Risk Premium}} = \beta_i \underbrace{(\boldsymbol{R_m - r_f})}_{\text{Market Factor}}$$

R_i	Return on Investment i
r_f	Risk-free Rate
B_i	Beta of Investment i
R_m	Market Return

- Note: Nothing (including asset own volatility) is relevant to the risk of an asset, except the beta of the asset (β_i), also called systematic risk.
- β_i>1: Aggressive Stocks and β_i<1: Defensive Stocks

Empirical Analysis

- The usual estimator of beta of the equity is the OLS estimator of the slope coefficient in the excess-return market model

$$\underbrace{\boldsymbol{R_{i,t}} - \boldsymbol{r_{f,t}}}_{\text{Stock Excess Return}} = \alpha_i + \beta_i \left(\underbrace{\boldsymbol{R_{m,t}} - \boldsymbol{r_{f,t}}}_{\text{Market Excess Return}}\right) + \epsilon_{i,t}$$

- α_i is the excess return (also known as the active return) above risk premium defined by CAPM in Sharpe (1964) and Lintner (1965).
- *Idiosyncratic Risk*$_i$ is residual sum of squares σ_ϵ

SSR

CAPM is not perfect

- Two fundamental features of CAPM has been rejected by data in thousands of empirical studies:
  - Asset risk premiums depend only on the asset's beta
  - There is only one factor that matters, the market portfolio.
- However, the basic intuition of the CAPM still holds true and it remains the workhorse model of finance:
  - 75% of finance professors advocate using it (Welch 2008), and
  - 75% of CFOs employ it in actual capital budgeting decisions (Graham and Harvey 2001)

CAPM Lessons

- **Risk is Factor Exposure:** The lower the correlation with a portfolio, the greater the diversification benefit
  - Low beta assets are associated with lower risk premium, since Investors do not need to be compensated much due to its diversification benefit.
- **Assets Paying Off In Bad Times Have Low Risk Premiums:**
  - Investors are, on average, risk averse so that the gains during good times do not cancel out the losses during bad times.
  - High beta assets are risky and require high expected returns to be held in equilibrium by investors.

Reasons for CAPM imperfection

- Many fundamental assumption of CAPM is too strong:
  - Investors have only financial wealth (No Labor Income)
  - Investors have mean-variance utility (Symmetric Treatment of Risk)
  - Investors have homogeneous expectations (Implicitly same belief)
  - Investors have no taxes, transactions costs, or need to rebalance
  - Investors (not as a whole) are price takers
  - Information is costless and available to all investors
- In summary, when the assumptions behind the CAPM are violated, we should expect additional risk premiums (in addition to market) should manifest themselves

Multifactor Models

- Multifactor models recognize that bad times (risk) can be defined more broadly than just bad returns on the market portfolio.

- The first multifactor model was the arbitrage pricing theory (APT), developed by Ross (1976). It uses the word “arbitrage” because the factors CANNOT be arbitraged or diversified away.

- In equilibrium, investors must be compensated for bearing multiple sources of factor risk (bad time).

Notes on Arbitrage Pricing Theory (APT)

- In APT model, active managers and arbitrageurs drive the expected return of assets toward a value consistent with an equilibrium trade-off between risk and return.
- The factors in the APT model are systematic ones, or those that affect the whole economy, that agents wish to hedge against.
- In addition, unlike the CAPM, the APT does not require the identification of the market portfolio. In fact, APT provides an approximate relation for expected asset returns with an unknown number of unidentified factors.

CAPM Revisit

- All assets have different exposures to the market factor and the greater the exposure, the higher the risk premium.

$$\underbrace{\boldsymbol{R_i - r_f}}_{\text{Risk Premium}} = \beta_i \, \underbrace{(\boldsymbol{R_m - r_f})}_{\text{Market Factor}}$$

- When the market does poorly, stocks that have high exposures to the market factor also tend to do badly. That is, high beta stocks tend to tank in parallel when the market tanks.

- Over the long run, the CAPM predicts that stocks with high betas will have higher average returns than the market portfolio to compensate investors for losses when bad times hit.

Risk Factors

- **Risk factors** offer **premiums** to compensate investors for bearing losses during bad times (risk)
- Fama & French (1993) model explains asset returns with three factors: **market factor** and **two additional factors**

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i\left(R_{m,t} - R_{f,t}\right) + s_i SMB_t + h_i HML_t + \varepsilon_{i,t}$$

  - SMB represents the risk associated with size
  - HML represents the risk associated with value

Size and Value factors

- *SMB*(**Small Minus Big**) capture the historical outperformance of small firms relative to large firms by the difference between the returns to portfolios of small- and large-capitalization firms

- *HML*(**High Minus Low**) the historical outperformance of value firms relative to value firms by the difference between the returns to portfolios of high and low Book Value/Market capitalization ratio firms

- Thus, the regression coefficients s_i and h_i represent exposures to size and value risk in much the same way that β_i measures the exposure to market risk

Interpretation

- SMB and HML factors are **factor mimicking portfolios**
  - Use CAPM concept of diversification to ensure that the factors capture size and value effects by averaging across many stocks
  - These factors are long-short portfolios and take positions away from the market portfolio
  - The average stock, however, only has market exposure (average investor holds the market) since every stock can't be small and every stock can't be large.

Note: Long-Short Portfolio

- LONG means BUY a selection of stocks (e.g., Value Stocks)
- SHORT means SHORT SELL the other (e.g., Growth Stocks)
  - Short selling is a trading strategy investors employ when they believe that shares of a stock or portfolio are overpriced.
  - It generally means borrowing a security or a basket of securities from a broker and selling it, with the understanding that it must later be bought back (hopefully at a lower price) and returned to the broker.
  - the short seller closes out the short position by purchasing equivalent securities on the open market

Size Factors in Reality

- The compound returns of SMB reach a maximum right around the early 1980 right after Banz and Reinganum studies were published

- Potential Interpretation:

  - The original discovery of the size premium could have just been "data mining"

    - Can not be repeated out of sample

  - The second response is that actually the size effect was there and actions of rational, active investors, acting on news of the finding, bid up the price of small cap stocks until the effect was removed.

Value Factors in Reality

- The book-to-market ratio is book value divided by market capitalization
  - In essence, a value strategy consists of buying stocks that have low prices (normalized by book value, sales, earnings, or dividends, etc.) and selling stocks that have high prices
- Academics often normalize by book value.
  - Value stocks are stocks with low prices relative to book value.
  - “Growth” stocks have high prices relative to book value.
- Practitioners often normalize prices by measures other than book value—which practitioners do when they build their (often proprietary) value factors.

Theoretical Interpretations

- **Rational Interpretation**: Value is Risky
  - Since not all risk can be diversified away, the remaining risk must be priced in equilibrium, leading to a value premium.
- **Behavioral Interpretation**: Investor Overreaction
  - The prices of growth firms are bid up too high irrationally. When this growth does not materialize, prices fall, leading to returns on growth stocks being low relative to value firms.

Definition of Alpha

- Alpha is the average return in excess of a benchmark

$$r_t^{ex} = r_t - r_t^{benchmark}$$

- The concept of alpha requires first defining a benchmark against which alpha can be measured.
- An ideal benchmarks would have the following characteristics:
  - Well Defined (Verifiable and ambiguity-free): Russell 1000
  - Tradable and Replicable

Factor Benchmarks

- Consider the CAPM applied to asset (or strategy or fund) *i*:

$$E(r_i) - r_f = \beta_i(E(r_m) - r_f)$$

- Let's say $\beta_i = 1.3 \;\rightarrow E(r_i) = 1.3E(r_m) - 0.3r_f$

- Benchmark of CAPM is a *Replicating portfolio*:
  - a short position in T-bills of -\$0.30 and a levered position in the market of \$1.30.

- The β_i is actually *a mimicking portfolio weight*

- Alpha of CAPM, or ***Risk-Adjusted Return*** (performance after risk considered):

$$\alpha_{i,t} = r_{i,t}^{ex} = r_{i,t} - \underbrace{E(r_{i,t})}_{\text{market risk factor}} = r_{i,t} - [1.3E(r_{m,t}) - 0.3r_f]$$

Factor Regression (CAPM)

- We can estimate the risk-adjusted benchmark, or equivalently the mimicking portfolio, using factor regressions
- To illustrate, consider the grand master of value investing, Warren Buffett. Taking 20-year monthly returns on Berkshire Hathaway at Dec 2018, we run the following CAPM regression:

$$E(r_i) - r_f = \underset{0.41\%}{\alpha_i} + \underset{0.44}{\beta_i}(E(r_m) - r_f)$$

- Buffett is generating an alpha of 0.0041 × 12 = 4.9% per year with a risk of less than half that of the market (ß = 0.44).

$1 investment is equivalent to

- 44 cents into market portfolio
- 56 cents into T-bill
- 0.41% (alpha) per month

Factor Regression (Fama French Factors)

- The SMB and HML factors are long-short factors. They are mimicking portfolios that consist of simultaneous \$1 long and \$1 short positions in different stocks:

$$SMB = \underbrace{\$1\ in\ small\ cap\ stocks}_{\text{Long}} - \underbrace{\$1\ in\ larg\ cap\ stocks}_{\text{Short}}$$

$$HML = \underbrace{\$1\ in\ value\ stocks}_{\text{Long}} - \underbrace{\$1\ in\ growth\ stocks}_{\text{Short}}$$

Factor Regression (Fama French Factors)

- The Fama–French benchmark can be estimated by running the following regression:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i\left(R_{m,t} - R_{f,t}\right) + s_i SMB_t + h_i HML_t + \varepsilon_{\ i,t}$$

- The SMB and HML factor loadings are given by s and h, respectively.
  - A stock starts moving with small stocks s_i becomes positive, and if it moves together with large stocks, s_i is negative.
  - Positive h_i indicate the stock has a value orientation, and negative h_i indicate the stock is acting like a growth stock.

Factor Regression (Fama French Factors)

- To illustrate, consider the grand master of value investing, Warren Buffett. Taking 20-year monthly returns on Berkshire Hathaway at Dec 2018, we run the following Fama French 3 factor regression:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i\left(R_{m,t} - R_{f,t}\right) + s_i SMB_t + h_i HML_t + \varepsilon_{i,t}$$

0.44% 0.58 -0.54 0.50

- If β_i stay the same only when the *SMB* and *HML* factors would have *no* ability to explain Buffett's returns.
- Negative *SMB* and positive *HML* factor loadings suggest that Berkshire Hathaway is a large, value investor.

Factor Regression (Fama French Factors)

- The surprising result in the Fama–French regression is that Buffett is still generating considerable profits relative to the size- and value-factor controls
  - Buffett's performance is clearly not merely from being a value investor, at least the way value is being measured relative to the CAPM.
- Investment Breakdown:

  $0.42 in T-bills + $0.58 in the market portfolio

  − $0.54 in small caps + $0.54 in large caps

  + $0.50 in value stocks − $0.50 in growth stocks

  + 0.44% (alpha) per month

Slide Deck #2 Supplementary:

Things you need to know

SELECT Query

Most basic query used to manipulate a database

SYNTAX:

```
SELECT column1, column2,
column3, … columnN
FROM table;
```

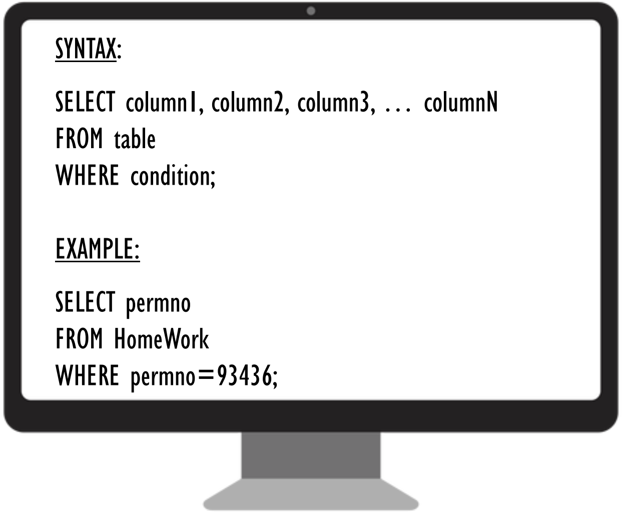
EXAMPLE:

```
SELECT permno FROM HomeWork;
SELECT * FROM HomeWork;
```

WHERE Query

If we need only certain records form a table.

where clause acts as Filtering mechanism.



AND

This operator returns a record if all the conditions separated by AND are True

SYNTAX:

```
SELECT column1, column2, column3, … columnN
FROM table
WHERE condition1 AND condition2;
```

EXAMPLE:

```
SELECT permno
FROM HomeWork
WHERE permno=93436 AND gvkey='184996';
```

OR

This operator returns a record if any of the conditions separated by OR are True

<u>SYNTAX</u>:

```
SELECT column1, column2, column3, … columnN
FROM table
WHERE condition1 OR condition2;
```

<u>EXAMPLE</u>:

```
SELECT permno
FROM HomeWork
WHERE permno=93436 OR permno=21936;
```

IN

This operator is used to specify multiple values inside the WHERE clause. It acts as a short for multiple OR

SYNTAX:

```
SELECT column1, column2, column3, …
FROM table
WHERE column1 IN (value1, value2, …);
```

EXAMPLE:

```
SELECT *
FROM HomeWork
WHERE permno IN (93436, 89393, 93002);
```

NOT

This operator returns a record if the condition(s) are NOT True

SYNTAX:

```
SELECT column1, column2, column3, … columnN
FROM table
WHERE NOT condition;
```

EXAMPLE:

```
SELECT permno
FROM HomeWork
WHERE NOT permno=93436;
```

COUNT Function

This function returns the number of rows that match specified criteria

SYNTAX:

```
SELECT COUNT(column)
FROM table
WHERE condition;
```

EXAMPLE:

```
SELECT COUNT(permno)
FROM HomeWork;
```

MIN(MAX) Function

This function returns the smallest (largest) value of the selected column

SYNTAX:

```
SELECT MAX(column)
FROM table
WHERE condition;
```

EXAMPLE:

```
SELECT MIN(permno)
FROM HomeWork;
```

AVERAGE Function

This function returns average value of a numeric column

SYNTAX:

```
SELECT AVE(column)
FROM table
WHERE condition;
```

EXAMPLE:

```
SELECT AVE(permno)
FROM HomeWork;
```

SUM Function

This function returns the total sum of a numeric column

SYNTAX:

```
SELECT SUM(column)
FROM table
WHERE condition;
```

EXAMPLE:

```
SELECT SUM(permno)
FROM HomeWork;
```

GROUP BY

Used in SQL to arrange identical data into groups with the help of some functions

<u>SYNTAX</u>:

```
SELECT Column;
FROM table
WHERE condition
GROUP BY Column(s);
```

<u>EXAMPLE:</u>

```
SELECT date, count(permno)
FROM HomeWork
GROUP BY date;
```

HAVING

Used to place conditions where we need to filter final returned results

<u>SYNTAX</u>:

```
SELECT Column;
FROM table
GROUP BY Column(s)
HAVING condition;
```

<u>EXAMPLE:</u>

```
SELECT date, count(permno)
FROM HomeWork
GROUP BY date
HAVING COUNT(permno)>1;
```

ORDER BY

This combination is used to sort the returned results in ascending or descending order

<u>SYNTAX</u>:

```
SELECT column1, column2, column3, …
FROM table
ORDER BY column1, column2, … (DESC);
```

<u>EXAMPLE:</u>

```
SELECT *
FROM HomeWork
ORDER BY date, permno;
```

ALIAS

Column ALIAS

<u>SYNTAX</u>:

```
SELECT column1 AS alias_name
FROM table;
```

<u>EXAMPLE:</u>

```
SELECT  permno AS stockID, gvkey AS CompanyID
FROM HomeWork;
```

ALIAS

Table ALIAS

SYNTAX:

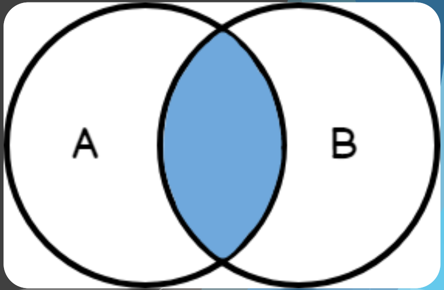
```
SELECT column1, column2, column3, …
FROM table AS Alias_name;
```

EXAMPLE:

```
SELECT  H.permno, H.CompanyName
FROM HomeWork as H;
```

INNER JOIN

This operation merges tables in the way below



SYNTAX:

```
SELECT a.column1, a.column2, b.column3, ...
FROM tableA AS a INNER JOIN tableB AS b
ON a.key=b.key;
```

EXAMPLE:

```
SELECT  H1.*
FROM HomeWork AS H1 INNER JOIN HomeWork AS H2
ON H1.date=H2.date;
```

SYNTAX:

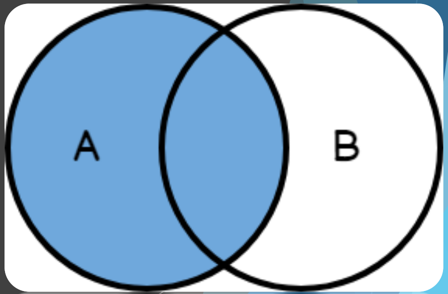
```
SELECT a.column1, a.column2, b.column3, …
FROM tableA AS a LEFT JOIN tableB AS b
ON a.key=b.key;
```

EXAMPLE:

```
SELECT H1.*
FROM HomeWork AS H1 LEFT JOIN HomeWork AS H2
ON H1.date=H2.date;
```

LEFT JOIN

This operation merges tables in the way below



CREATE TABLE & DROP TABLE

Operations to create (drop) tables in a database (a composite of many tables)

<u>SYNTAX:</u>

```
CREATE TABLE newtable as
SELECT a.column1, a.column2, b.column3, ...
FROM tableA AS a LEFT JOIN tableB AS b
ON a.key=b.key;
```

<u>EXAMPLE:</u>

```
CREATE TABLE HW1 as
SELECT  H1.*
FROM HomeWork AS H1;
```

<u>EXAMPLE:</u>

```
DROP TABLE HW1;
```

Cumulative Return (a SQL trick)

- Transformation: compounding cumulative return over a holding period

$$R_H = \frac{P_{t+H}}{P_t} - 1 \rightarrow 1 + R_H = \frac{P_{t+H}}{P_t}$$

1. $1 + R_H = \frac{P_{t+1}}{P_t} \times \frac{P_{t+2}}{P_{t+1}} \times \cdots \times \frac{P_{t+H}}{P_{t+H-1}}$
2. $1 + R_H = (1 + R_{t+1}) \times (1 + R_{t+2}) \times \cdots \times (1 + R_{t+H})$
3. $1 + R_H = e^{log(1+R_{t+1})} \times e^{log(1+R_{t+2})} \times \cdots \times e^{\log(1+R_{t+H})}$
4. $1 + R_H = e^{\sum_i log(1+R_{t+i})}$
5. $R_H = e^{\sum_i log(1+R_{t+i})} - 1$

Note: when there were dividends, we assume we reinvest the cash amount.

Cumulative Return (a SQL trick)

```
proc sql;
    create table cumRet as
    select distinct gvkey, permno, fyear,
    exp(sum(log(1+ret)))-1 as cumRet,
    count(distinct date) as months,
    max(distinct ret) as max_ret
    from ccm_matched
    group by gvkey, permno, fyear;
quit;
```

Slide Deck #3:
More Factors

- A Literature Review

Table 6

- Regression of excess stock and bond returns (in percent) on the excess market return (R_m-R_f) and the mincing return for the size (*SMB*) and book-to-market equity (*HML*) factor July 1963 to December 1991, 342 months.

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + s_i SMB_t + h_i HML_t + \varepsilon_{i,t}$$

- $R_m$is the value-weighted percent monthly return on all the stocks in the 25 size-BE/ME portfolios, plus the negative-BE stocks excluded from the 25 portfolios.
- R_f is the one-month Treasury bill rate, observed at the beginning of the month.

Empirical Standard

- Carhart (1997) and contemporaneous papers establish an standard way to demonstrate a new factor
- New Factors is defined the difference between top and bottom portfolios
  - Long the Top One
  - Short the Bottom One
- Empirical Demonstration
  1. Monotonic increase (decrease) of CAPM or FF alphas from bottom to top portfolios
  2. The Alpha from Long (one extreme) and Short (the other extreme) portfolio is statistically and economically significant
  3. A significance of the newly found Alpha cannot be explained by other known factors

The Fama French + Momentum Model

- Momentum is the strategy of buying stocks that have gone up over the past six (or so) months (winners) and shorting stocks with the lowest returns over the same period (losers).
  - The momentum effect refers to the phenomenon that winner stocks continue to win and losers continue to lose.

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i\left(R_{m,t} - R_{f,t}\right) + s_i SMB_t + h_i HML_t + u_i UMD_t + \varepsilon_{i,t}$$

- The new mimicking factor is called UMD, for stocks that have gone up minus stocks that have gone down.
- Winners and losers are always relative—stocks win or lose relative to each other, and the market as a whole can go up or down.

Momentum Calculation

- Fama French's Formula:
- Similar to B/M factor, 6 value-weight portfolios formed on size and prior (2-12) returns to construct UMD.
  - the intersections of 2 portfolios formed on size (market equity, ME) and 3 portfolios formed on prior (2-12) return.
  - The monthly size breakpoint is the median NYSE market equity.
  - The monthly prior (2-12) return breakpoints are the 30th and 70th NYSE percentiles.

$$UMD_t = AVG\left(R_{t,Port_{S\&U}}, R_{t,Port_{B\&U}}\right) - AVG\left(R_{t,Port_{S\&D}}, R_{t,Port_{B\&D}}\right)$$

Momentum Risk

- Momentum returns are not the opposite of value returns (the correlation of HML with UMD is –18% over 1965 to 2018).
  ±30% caution line
  - many investors who claim that they are growth investors are actually momentum investors
- There is one sense in which momentum is the opposite of value.
  - Value is a negative feedback strategy: A stabilizing strategy
  - Momentum is a positive feedback strategy: destabilizing Strategy (subject to periodic crashes)

周期性投资

Volatility Sensitivities

- Different sensitivities to aggregate volatility innovations (proxied by ΔVIX) have different average returns

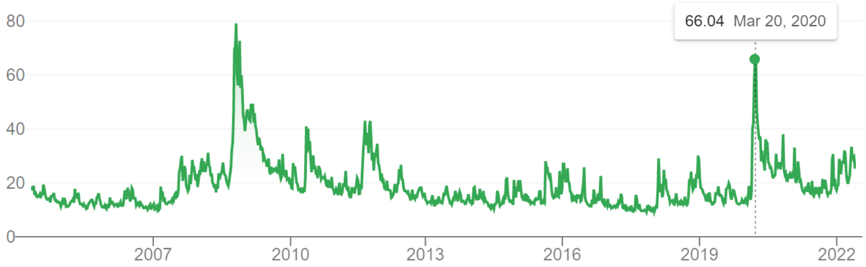
  - VIX represent equity market volatility - a popular measure of the stock market's expectation of volatility implied by S&P 500 index options.

$$\boldsymbol{R_{i,t}} - \boldsymbol{r_{f,t}} = \beta_i\,(\boldsymbol{R_{m,t}} - \boldsymbol{r_f}) + \beta_{\Delta VIX,i}\Delta VIX_{\boldsymbol{t}}$$

- We form value-weighted quintile portfolios every month by regressing excess individual stock returns on ΔVIX, controlling for the MKT factor using daily data over the previous month.

Notes on VIX Index

- VIX is the ticker symbol for the Chicago Board Options Exchange's Volatility Index, a popular gauge of the stock market's expectation of volatility based on S&P 500 index options.
- The VIX index, calculated using options-based theory and current options-market data, serves as a proxy for market volatility in the near future (30 days).
- The CBOE calculates and disseminates it in real time, and it is also known as the fear index or fear gauge.



Anomaly and Firm Characteristic Factors

- Fama and French (1996) calls firm/stock characteristics that are related to average returns on common stocks and cannot be explained by CAPM as *Anomalies*. 异常现象

- A risk factor should be a variable that has unpredictable variation through time. Assets' risk exposures to this factor need to be able to explain the cross-sectional return patterns

- Certain firm characteristic is found to be correlated with the cross-section of expected returns, ***a long-short portfolio*** can usually be constructed to proxy for the underlying unknown risk factor.

Idiosyncratic and Total Volatility

- Total volatility is standard deviation of stock return

$$Total\ Volatility = \sqrt{var(R_{i,t})}$$

- Idiosyncratic Volatility is measured by following regression

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i\left(R_{m,t} - R_{f,t}\right) + s_i SMB_t + h_i HML_t + \varepsilon_{\ i,t}$$

  - In particular,

22 trading days

$$Idiosyncratic\ Volatility = \sqrt{var(\varepsilon_{i,t})}$$

Daily

  - Monthly formation of value-weighted quintile portfolios by sorting stocks based on total volatility and idiosyncratic volatility relative to the Fama-French (1993) model, using daily data over the previous month.

Volatility (low-risk) Anomaly

- The whole point of the CAPM and the many multifactor extensions was that stock return volatility itself should not matter. Expected returns are determined by how assets covary with factor risks.
  - Idiosyncratic volatility should definitely not have any relation to expected returns under the CAPM.
- In theory, higher idiosyncratic risk would reduce demand for a stock. Hence, agents have to be compensated for bearing idiosyncratic risk, i.e. lower price than otherwise similar stocks, resulting in a positive relation between idiosyncratic risk and cross-sectional return in equilibrium.
- However, Ang et al. (2006) results show exactly the opposite.

Volatility (low-risk) Anomaly

- The robustness of the negative relationship between idiosyncratic and overall volatility and returns is particularly noteworthy.
- Ang et al. (2006) employed a large number of controls for size, value, leverage, liquidity risk, volume, turnover, bid–ask spreads, co-skewness risk, dispersion in analysts' forecasts, and momentum.
- In subsequent work, Ang et al. (2009) showed that the volatility effect existed in each G7 country and across all developed stock markets.
  - They also added controls for private information, transactions costs, analyst coverage, institutional ownership, and delay measures.

Leveraged Portfolio?

- Frazzini and Pedersen (2010) construct a betting-against-beta (BAB) factor that goes long low-beta stocks and short high-beta stocks.
- As the differences in average returns across the beta quintiles are small, the authors form their BAB factor by scaling the low- and high-beta portfolios by betas (only two portfolios both of which are beta-weighted and initially has a unit-beta), as low beta can be close to zero:

$$r_{t+1}^{BAB} = \frac{1}{\beta_t^L}(r_{t+1}^L - r^f) - \frac{1}{\beta_t^H}(r_{t+1}^H - r^f)$$

- The BAB is the self-financing zero-beta portfolio. For example, on average, the US stock BAB factor is long \$1.4 of low-beta stocks (financed by shorting \$1.4 of risk-free securities) and short sells \$0.7 of high-beta stocks (with \$0.7 earning the risk-free rate).

Systematic Risk

同时发生的

- The CAPM states a contemporaneous relation between beta and expected returns, i.e., no explicit statement on lagged betas.
  - Positive contemporaneous beta and return relationship (which is true, though may be weak)
  - BAB: Negative relationship between lagged beta and return

- Fama and French (1992) noted that “beta shows no power to explain average returns”, while their main results showed that size and value effects dominated beta in individual stocks.

Gambling Investment Behavior

- Motivated by existing evidence of a preference among investors for assets with lottery-like payoffs and that many investors are poorly diversified; the authors investigate the significance of extreme positive returns in the cross-sectional pricing of stocks.
- Authors sort stocks by their maximum daily return during the previous month and examine the monthly returns on the resulting portfolios over the period July 1962–December 2005.
- This evidence suggests that investors may be willing to pay more for stocks that exhibit extreme positive returns, and thus, these stocks exhibit lower returns in the future.

A Working Growth Strategy

- Fama and French (2008) argues that "profitability... do not provide much basis for the conclusion that, with controls for market cap and B/M, there is a positive relation between average returns and profitability."
- However, Novy-Marx (2013) found that profitability, measured by gross profits-to-assets, [(REVT − COGS)/AT], has roughly the same power as book-to-market predicting the cross section of average returns.

  Revenue
- Strategies based on gross profitability generate value-like average excess returns, even though they are growth strategies that provide an excellent hedge for value.

The Q Factors

- A similar Setting to Fama French (1993):
  - $E[r^i] - r^f = \beta^i_{MKT}E[MKT] + \beta^i_{ME}E[r_{ME}] + \beta^i_{I/A}E[r_{I/A}] + \beta^i_{ROE}E[r_{ROE}]$
    - $E[MKT]$ is the excess return defined by CAPM
    - $E[r_{ME}]$ is similar to Fama and French SMB
    - $E[r_{I/A}]$ is the difference between the return on a portfolio of low investment stocks ($\Delta AT_t/AT_{t-1}$) and the return on a portfolio of high investment stocks
    - $E[r_{ROE}]$ is the difference between the return on a portfolio of high profitability (return on equity, ROE) stocks and the return on a portfolio of low profitability stocks

- However, HML is dropped, because
  - The investment factor has a high correlation of 0.69 with HML, and the ROE factor has a high correlation of 0.50 with UMD
  - The alphas of HML and UMD in the q-factor model are small and insignificant, but the alphas of the investment and ROE factors in the Carhart model are large and significant.

Fama and French’s Reply

- Fama French 5 Factors:
  - $R_{i,t} - R_{f,t} = \alpha_i + \beta_i\left(R_{m,t} - R_{f,t}\right) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + \varepsilon_{\,i,t}$
    - RMW_t is the difference between the returns on diversified portfolios of stocks with robust and weak profitability
      - Profitability of June of year t is defined as annual revenues minus cost of goods sold, interest expense, and selling, general, and administrative expenses divided by book equity for the last fiscal year end in t-1
    - CMA_t is the difference between the returns on diversified portfolios of the stocks of low and high investment firms, which we call conservative and aggressive.
      - Investment is defined as the change in total assets from the fiscal year ending in year t-2 to the fiscal year ending in t-1, divided by t-2 total assets at the end of each June
  - Also a theoretical argument (Miller and Modigliani 1961) is brought in into this 5-factor paper.
  - A follow-up RFS paper shows the 5-Factor Model helps explain the low average stock returns associated with high B, large share issues, and highly volatile returns.

ELSEVIER

Journal of Financial Economics

Volume 116, Issue 1, April 2015, Pages 1-22

A five-factor asset pricing model ☆

Eugene F. Fama [a], Kenneth R. French [b]

Show more

https://doi.org/10.1016/j.jfineco.2014.10.010

Abstract

A five-factor model directed at capturing the size, value, profitability, and investment patterns in average stock returns performs better than the three-factor model of Fama and French (FF, 1993). The five-factor model's main problem is its failure to capture the low average returns on small stocks whose returns behave like those of firms that invest a lot despite low profitability. The model's performance is not sensitive to the way its factors are defined. With the addition of profitability and investment factors, the value factor of the FF three-factor model becomes redundant for describing average returns in the sample we examine.

What is this paper about?

- At least 316 factors have been tested to explain the cross-section of expected returns.

- Given the inevitable data mining, many of the historically discovered factors would be deemed “significant” by chance.

- The authors suggest that a newly discovered factor today should have a *t*-statistic that exceeds 3.

Surprising results and what next?

- Standard predictive regressions fail to reject the hypothesis the followings are significantly related to factors:
  - party of the US President
  - global warming
  - the conjunctions of the planets
  - …
- The author urge us to be careful to reject those seemingly irrelevant results, as doing do so entails rejecting the standard methodology on which the return predictability literature is built.
- Instead, the success of the tests performed here could encourage further research with more care.

Back to 80’s

- Merton, R.C., 1980. On Estimating the expected return to the market: an exploratory investigation. Journal of Financial Economics
  - “Even if the expected return...were known to be a constant for all time, it would take a very long history of returns to obtain an accurate estimate.”
  - “difficult to use the time series of realized returns to distinguish among different models for expected return.”

Robert C. Merton

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1997

Prize motivation: "for a new method to determine the value of derivatives"
Born: 31 July 1944, New York, NY, USA

Slide Deck #3 Supplementary:

Things you need to know

Calculating Portfolio Returns

起初

- Equal Weighting assigns an equal weight to each constituent stock at inception and each rebalance:
  - This index is calculated as the arithmetic average return
  - biased towards small stocks

- Market capitalization (Value) weighting assigns each constituent stock a weight determined by the stock's market capitalization proportion in the stock universe.
  - The weighting would not be constant, as long as there are price changes
  - Biased towards big stocks (and resulting momentum tilt)

倾斜

- Fundamental (factor) weighting gives portfolio a factor title by weighting constituent stocks based on certain factors, such as B/M, Size, and Dividends.

Equal weighted example

交易挂单价格下降，有损失

Need to rebalance every month

In our case, it is previous end close price

Simply, $\frac{P_t+D_{t-1,t}-P_{t-1}}{P_{t-1}}$

Security	Month Beginning Price	Weight	Jun 30 Value	Holding Shares	Month End Price (close)	Dividend Paid	July 31 Value	Return
A	40	20.00%	$2,000	50	44	0	$2,200	10.00%
B	10	20.00%	$2,000	200	14	0	$2,800	40.00%
C	4	20.00%	$2,000	500	5	0	$2,500	25.00%
D	12.5	20.00%	$2,000	160	7	2	$1,440	-28.00%
E	25	20.00%	$2,000	80	20	0	$1,600	-20.00%
Total			$10,000				$10,540	5.40%

Ret=0.2(10%+40%+25%-28%-20%)
or Ret=(10540-10000)/10000

Value weighted example

BTW, you can assign weights according to firm characteristics (e.g., B/M or IVOL) in a similar way.

Less need to rebalance every month in our case
- other than reinvest dividends

Security	Month Beginning Price	Weight	Jun 30 Value	Holding Shares	Month End Price (close)	Dividend Paid	July 31 Value	Return
A	40	25.00%	$2,500	63	44	0	$2,772	10.00%
B	10	11.00%	$1,100	110	14	0	$1,540	40.00%
C	4	13.00%	$1,300	325	5	0	$1,625	25.00%
D	12.5	9.00%	$900	72	7	2	$648	-28.00%
E	25	42.00%	$4,200	168	20	0	$3,360	-20.00%
Total			$10,000	$10,020			$9,945	-0.77%

Ret= (25%*10%+11%*40%+13%*25%-9%*28%-42*20%) =-0.77%

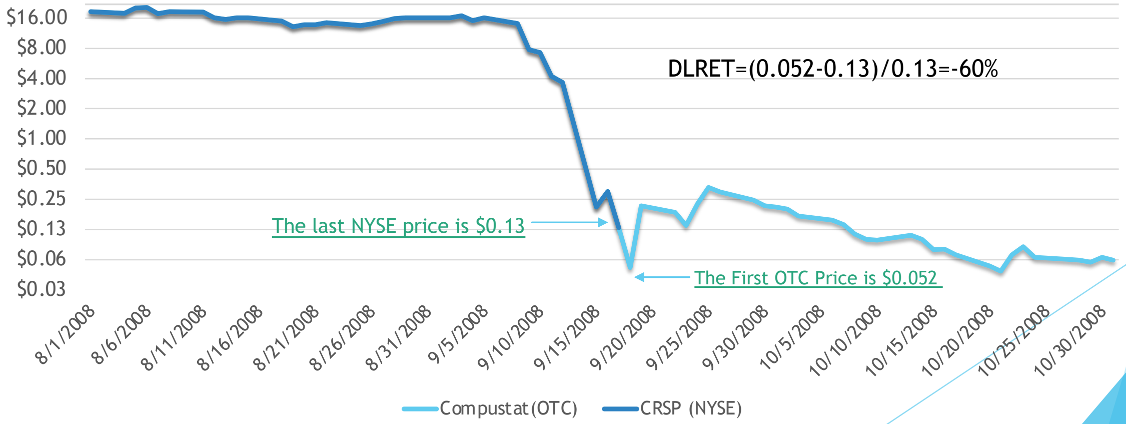
… Ret=(9945-10020)/10020=-0.75%

Back

Delisting Return

- Delisting returns are computed from total value of distributions to shareholders (liquidation payments) or from a price on another exchange.

- Lehman Brothers' Case



About CRSP Monthly Return

- CRSP monthly return does not count partial returns. The partial monthly return is considered as a part of delisting return:
- In Lehman Brothers cases:
  - Its last non-empty monthly return is in August 2008
  - Its return value in Sep 2008 is empty, and its delisting return (at Sep 2008) is

$$DLRET = \frac{Value\ after\ Delisting}{Price\ at\ Last\ Trading\ Day\ of\ Previous\ Month} - 1 = \frac{\$0.052}{\$16.09} - 1 = -99.68\%$$

  - It would be more appropriate to include delisting return for certain research applications

Back

Delisting Return, Cont'd

- Based on Beaver, et al (2007):
  - The delisting events primarily include Mergers and Dropped delistings
    - Dropped delistings include bankruptcy, stock price below acceptable level, and insufficient assets, equity, or capital

Table 2 -Panel A: *sample size with earnings$_{t+1}$ (1962 to 2002)*

Firm-Years	With Earning$_{t+1}$	Without Earning$_{t+1}$
Delisting		
Merger-related (200–299)	774	4,803
Dropped (500–599)	2,849	1,970
Other	146	378
Total Delisting Obs	**3,769**	**7,151**
Total CRSP and Compustat firm-year Obs	**146,082**	**7,887**

Coding Notes

- You can treat each delisting return DLRET as an additional RET like a value in a date range for which cumulative returns are calculated. It would not be added to RET, but treated as a multiplicative factor to get the exact effect.
  - Namely, (1+DLRET)*(1+RET)-1 on the last trading date can serve as a fair approximation.
- Note the "date" variable for the delisting events from CRSP event data is not necessarily the day the stock is liquidated. According to Beaver, et al (2007),
  - 79% of delisting distribution payments are made in the month of the delisting
  - 16% are made after the delisting month and within three months of the delisting.
  - The remaining payments occur more than three months after the delisting month.

Back

Slide Deck #4:

Fama & French 1993 replication

A Perfect World (CAPM)

- Capital Asset Pricing Model is a model based on equilibrium with agents having mean-variance preferences:
$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i\left(R_{m,t} - R_{f,t}\right) + \varepsilon_{\ i,t}$$

- Implication: A model measuring “normal” returns
  - Expected returns of a stock are a linear function of their <u>Market βs</u>
$$E\left[R_{i,t} - R_{f,t}\right] = E\left[R_{m,t} - R_{f,t}\right] \times \beta_i$$
  - <u>Market βs</u> is sufficient to describe the cross-section of expected stock returns
  - α_i (Jensen’s) captures the return component cannot be theoretically explained

Anomalies and Market Efficiency

- Assuming CAPM is adequate to reflect the market efficiency, average stock returns should not be related to firm characteristics

- Well-known anomalies:
  - Banz (1981) showed that small-capitalization firms are associated with higher average returns than is predicted CAPM (*SIZE*)
  - Basu (1977) noted that firms with high E/P earn positive abnormal returns relative to the CAPM
  - Subsequently, Stattman (1980) find average stock returns are positively related to B/M (*VALUE*)
  - And many more…

Fama and French 3 Factor Models

- Fama and French (1992, 1993, and 1996) have argued that those anomalies are captured due to the inadequacy of CAPM, (rather than market inefficiency).

- They argue that CAPM fails to capture risk factors, such as SIZE and VALUE

- They suggest following empirical model to measure abnormal performance (α_i):

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i\left(R_{m,t} - R_{f,t}\right) + s_i SMB_t + h_i HML_t + \varepsilon_{i,t}$$

- In this equation, *SMB* (*HML*) represent the risk associated with ***SIZE*** (***VALUE***).

New Risk Factors

- *SMB*(**S**mall **M**inus **B**ig) represents the difference between the returns to portfolios of small- and large-capitalization firms
  - A portfolio return from buying stocks of small firms while selling stocks of large firms
- *HML*(**H**igh **M**inus **L**ow) represents the difference between the returns to portfolios of high and low B/M (Book Value/Market capitalization) ratio firms
  - A portfolio return from buying stocks of value firms while selling stocks of growth firms
- Thus, the regression coefficients s_i and h_i represent exposures to size and value risk in much the same way that β_i measures the exposure to market risk

Factor Constructions

- To capture the SIZE and VALUE anomalies
  - “The splits are arbitrary,… and we have not searched over alternatives”
  - Portfolio returns are all value-weighted portfolio returns

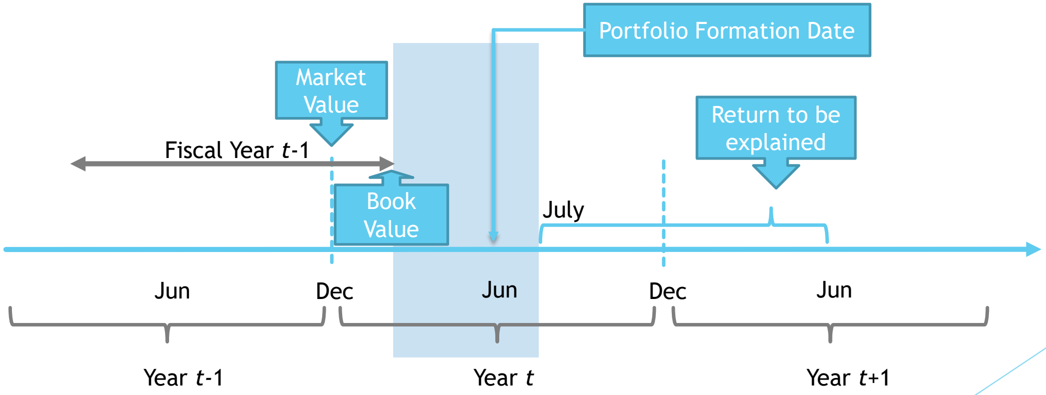
	Small	Big
High B/M (Value)	S&H	B&H
Middle B/M (Neutral)	S&M	B&M
Low B/M (Grow)	S&L	B&L

- $HML_t = AVG\left(R_{t,Port_{S\&H}}, R_{t,Port_{B\&H}}\right) - AVG\left(R_{t,Port_{S\&L}}, R_{t,Port_{B\&L}}\right)$
- <holding constant the SIZE ratios>

- $SMB_t = AVG\left(R_{t,Port_{S\&H}}, R_{t,Port_{S\&M}}, R_{t,Port_{S\&L}}\right) - AVG\left(R_{t,Port_{B\&H}}, R_{t,Port_{B\&M}}, R_{t,Port_{B\&L}}\right)$
- <holding constant the B/M ratios>

Available Information

- All relevant information must be known before the returns they are used to explain:



CRSP Prep

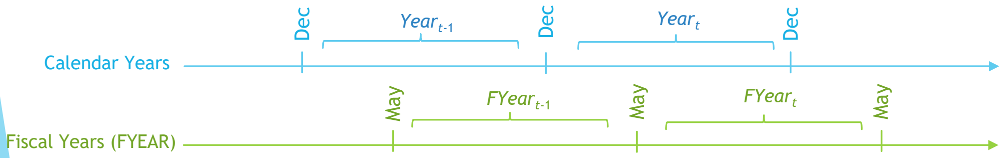
- A firm must have stocks prices for Dec at year *t*-1, and Jun at year *t*
- Keep stocks with share code of 10 or 11 (Exclude ADRs, ETFs, etc.)
- Back to 1962: US stocks are integrated progressively
    - NYSE at 1925, NYSE MKT(AMEX) in 1962, Nasdaq in 1972, ARCA in 2006, and Bats in 2012
        - Stocks listed in NYSE MKT and Nasdaq also tended to be smaller than NYSE stocks
        - Note: Bats are backfilled around 2018, which retrospectively affected CRSP equal and value weighted index back to 2012
    - NYSE breakpoints are always used to ensure consistence and representativeness
        - "The median NYSE size is then used to split NYSE, Amex. and NASDAQ stocks into two groups, small and big (S and B)."

Compustat Prep

- A firm must have positive book common equity for year *t*-1 (Fiscal)
- book common equity is measured as
  - stockholders' equity + balance-sheet deferred taxes - book value of preferred stock
  - 赎回 清算
    depending on availability, the redemption, liquidation, or par value (in that order) is used to estimate the value of preferred stock
- A firm must have appeared in Compustat for two years
  - Survival bias: In early years, Compustat only includes firms with more than two years of historical data when it adds firms (i.e., companies that did not make its third year are excluded)

Final Empirical Notes

- CRSP and Compustat universe are different, but neither is a superset of the other
- Portfolio Construction is at each Jun, i.e., the portfolio is rebalanced or reconstructed at the last trading date of each June
  - *SMB* is based on the Market Capitalization at each June end
  - *HML* is, however, based on
    1. Book value of equity at the end of Fiscal year *t*-1 (previous Jun to current May), and
    2. Market value of equity at the end (December) of calendar year *t*-1



Compustat Data Handling

- *"We define book common equity, BE as the COMPUSTAT book value of stockholders equity, plus balance-sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock."*
  - *"Depending on availability, we use the redemption, liquidation, or par value (in that order) to estimate the value of preferred stock."*

```
proc sql;
            create table _comp_ as
            select distinct gvkey, fyear, datadate,
            coalesce(PSTKRV,PSTKL,PSTK,0) as ps,  /* Formula used by FF(1993) */
            seq + sum(TXDITC,0) - calculated ps as be 'Book Value', /* Formula used by FF(1993) */
            year(datadate) as year /*Calendar Year*/
            from comp.funda
            where indfmt='INDL' and datafmt='STD' and popsrc='D' and consol='C' /*"Standard" Compustat Filters*/;

            create table comp as
            select distinct *, fyear-min(fyear) as age /*Checking existing years*/
            from _comp_
            group by gvkey;
quit;
```

❖ Note Inclusion of Investment Tax Credit (ITCB) would reduce the correlation coefficient of SML and HTML outputs of this code and the ones from Professor Kenneth R. French's Data Library.

CRSP Data Handling

- The Fama and French (1993) and many follow up papers do not adjust return of equity with delisted return, which would be discussed more in the Event Study Notes.
- The delisted return treatment, however, is more appropriate for construction of mimicking portfolio as stock delisting is rarely a foreseeable event.

```
proc sql;
            create table msf as
            select distinct a.permco, a.permno, a.date, prc, ret, shrout, retx, shrcd, exchcd
            from crsp.msf(where=(date>='01Dec1961'd)) as a inner join
            crsp.msenames(where=(exchcd in (1,2,3))) as b
            /*1)NYSE(exchcd=1), AMEX(=2), and NASDQA (=3)*/
            on a.permno=b.permno and NAMEDT<=date<=NAMEENDT;

            create table crspm as
            select a.*, b.dlret, sum(1,ret)*sum(1,dlret)-1 as retadj
            "Return adjusted for delisting"/*2)Adding Delisted Returns*/,
            abs(a.prc)*a.shrout as meq 'Market Value of Equity at issue level'
            from msf as a left join crsp.msedelist(where=(missing(dlret)=0)) as b
            on a.permno=b.permno and
            intnx('month',a.date,0, 'END')=intnx('month',b.DLSTDT,0,'END')
            order by a.date, a.permco, MEq;
quit;
```

CRSP Data – Market Value of Equity

```
proc sql;
          create table crspm as
          select distinct *, sum(meq) as me
          from crspm
          group by permco, date;
quit;
```

TICKER	PERMCO	PERMNO	DATE	MEQ	ME
GOOGL	45483	90319	20180531	328,521,600	707,131,027
GOOG	45483	14542	20180531	378,609,427	707,131,027
GOOGL	45483	90319	20180629	337,239,351	726,547,658
GOOG	45483	14542	20180629	389,308,307	726,547,658

- Note one way to determine primary shares is to use the shares with larger Market Cap, though this approach many not be always right. Stock with ticker GOOG is, for example, Class-C shares (shares without voting rights)

Code

CRSP Data –Rebalance

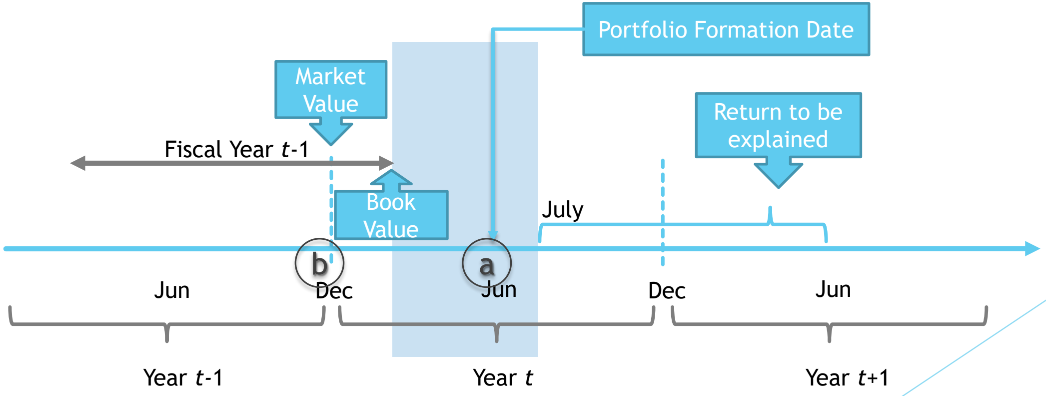
- The code below is conceptually the most straightforward solution to construct a value-weighted portfolio
  - All dividends are assumed to be reinvested proportionally based on the market value of stocks at the previous month end.
  - The code appendix provides another approach suggested by Professor Robert F. Stambaugh at Wharton School, which may further improve the correlation coefficients of calculated SMB and HML with the ones from Professor Kenneth R. French’s Data Library.

```
proc sql;
      create table crspm_w as
      select distinct a.permno, a.date, a.retadj, a.me, a.exchcd, a.shrcd, b.meq as weight_port
      from crspm as a inner join crspm as b
      on a.permno=b.permno and intnx('month',a.date,-1,'E')=intnx('month',b.date,0,'E');
quit;
```

Code

CRSP Data - Size and B/M Preparation

```
proc sql;
        create table crspjune as
        select a.permno, a.date, a.exchcd, a.shrcd, a.me as JUN_ME "Market Cap in current June"/*a*/,
        b.me as DEC_ME "Market Cap in previous December"/*b*/
        from crspm(where=(month(date)=6)) as a inner join
        (select distinct * from crspm(where=(month(date)=12))) as b
        on a.permno=b.permno and intck('month',b.date,a.date)=6;
quit;
```



CRSP & COMPUSTAT Matching

```
proc sql;
            create table ccm as
            select a.*, b.lpermno as permno, b.linkprim
            from comp as a, crsp.Ccmxpf_lnkhist as b
            where a.gvkey=b.gvkey
            and linktype in ('LC' 'LU') /*1)Confirmed Linkage Only*/
            and linkprim in ('P','C') /*2)Take Primary Shres Only*/
            and linkdt<=datadate<=coalesce(linkenddt, today())
            /*3)Matched based on the first valid June after Fiscal Year END*/
            group by permno, fyear
            /*4)Keep the Last Datadate for a Calendar Year^*/
            having datadate=max(datadate)
            /*5)Remove Duplicated Records through disagreed primary shares*/
            order by permno, year, datadate;;
quit;
```

Potential duplications:

- A few cases contain different GVKEYS for a same PERMNO-Date combination
- Some companies change length of fiscal year end

- For Matching CRSP and Compustat with CUSIP, please see Case Study at the end of this class.

CCM - Size and B/M Calculation

- *"[T]o avoid the survival bias inherent in the way COMPUSTAT adds firms to its tapes, we do not include firms until they have appeared on COMPUSTAT for two years."*
    - Adding this condition would slightly reduce the correlation coefficient, which may imply that the official "Fama French" factors may no longer take consideration of the survivorship bias documented in early papers.
    - In fact, Compustat now updates (and corrects) its data on daily basis, so that this previously documented bias may no longer exist in recent Compustat data.

```
proc sql;
            create table ccm_june as
            select a.*, b.BE, (1000*b.BE)/a.DEC_ME as BEME, b.age, b.datadate
            from crspjune as a, ccm as b
            where a.permno=b.permno and age > 2  and be > . and
            intnx('month',a.date,-12,'E')<datadate<=intnx('month',a.date,0,'E')
            order by a.date;
quit;
```

Code

NYSE Break Point Calculation

```
proc univariate data=ccm_june noprint;
            where exchcd=1 and beme>0 and shrcd in (10,11) /*Common US Stocks*/;
            var JUN_ME BEME;
            by date;
            output out=nyse_breaks median = SIZEMEDN
                        pctlpre=JUN_ME BEME pctlpts=30 70;
run;

proc sql;
            create table ccm_nyse_port as
            select a.permno, a.date, a.exchcd, a.shrcd,
            case when JUN_ME <= sizemedn then 'S' else 'B' end as sizeport,
            case when beme <= beme30 then 'L' when beme > beme70 then 'H'
            else 'M' end as btmport
            from ccm_june as a inner join nyse_breaks as b
            on a.date=b.date and shrcd in (10,11);
quit;
```

	Small	Big
High B/M (Value)	S&H	B&H
Middle B/M (Neutral)	S&M	B&M
Low B/M (Grow)	S&L	B&L

❖In the original Fama & French (1993), size and B/M break breakpoints are created by NYSE stocks only to ensure consistency introduced by the small stocks listed in the later joined NASDAQ (1973).

❖This treatment becomes a standard way to handle breaking points among Asset Pricing literature, even in the ones using data after CRSP NASDAQ data integration.

Code

NYSE Break Portfolio Construction

```
proc sql;
            create table port_ccm as
            select a.*, b.sizeport, b.btmport, b.date as portdate format date9.
            from crspm_w as a, ccm_nyse_port as b
            where a.permno=b.permno and  1 <= intck('month',b.date, a.date) <= 12
            /*1) Match Previous June to current July to June*/
            order by date, sizeport, btmport;

            create table port_ccm_vwret as
            select distinct date, sizeport, btmport, count(distinct permno) as n_firms,
            sum(weight_port*retadj)/sum(weight_port) as vwret
            from port_ccm
            group by date, sizeport, btmport;
quit;

proc transpose data=port_ccm_vwret(keep=date sizeport btmport vwret)
            out=ff_vwret (drop=_name_);
            by date ;
            ID sizeport btmport;
            Var vwret;
run;
```

❖ $Value\ Weighted\ Return_t = \sum_i^N (w_{i,t-1} \times R_t) = \sum_i^N \left(\frac{Market\ Cap\ of\ Stock_{i,t-1}}{\sum_i^N (Market\ Cap\ of\ Stock_{i,t-1})} \times R_t \right) = \frac{\sum_i^N (Market\ Cap\ of\ Stock_{i,t-1} \times R_t)}{\sum_i^N (Market\ Cap\ of\ Stock_{i,t-1})}$

A Constant

Fama & French Factors

```
data ff_factors; set ff_vwret;
          WH = (bh + sh)/2;
          WL = (sl + bl)/2;
          WHML = WH - WL;
          WB = (bl + bm + bh)/3 ;
          WS = (sl + sm + sh)/3 ;
          WSMB = WS - WB;
          label WH   = 'WRDS High'
                WL   = 'WRDS Low'
                WHML = 'WRDS HML'
                WS   = 'WRDS Small'
                WB   = 'WRDS Big'
                WSMB = 'WRDS SMB';
run;
```

	Small	Big
High B/M (Value)	S&H	B&H
Middle B/M (Neutral)	S&M	B&M
Low B/M (Grow)	S&L	B&L

- $HML_t = AVG\left(R_{t,Port_{S\&H}}, R_{t,Port_{B\&H}}\right) - AVG\left(R_{t,Port_{S\&L}}, R_{t,Port_{B\&L}}\right)$
- $SMB_t = AVG\left(R_{t,Port_{S\&H}}, R_{t,Port_{S\&M}}, R_{t,Port_{S\&L}}\right) - AVG\left(R_{t,Port_{B\&H}}, R_{t,Port_{B\&M}}, R_{t,Port_{B\&L}}\right)$

Result Comparison

```
proc sql;
        create table ff_test as
        select distinct a.*, smb, hml
        from ff_factors as a inner join ff.factors_monthly as b
        on intnx('month', a.date, 0, 'e')=intnx('month', b.date, 0, 'e');
quit;

proc corr data=ff_test;
        var smb hml;
        with wsmb whml;
run;

proc corr data=ff_test(where=(year(date)>1980));
        var smb hml;
        with wsmb whml;
run;
```

Pearson Correlation Coefficients, N = 690 Prob > |r| under H0: Rho=0		
	smb	hml
WSMB	0.98461 <.0001	-0.13715 <.0001
WHML	-0.18290 <.0001	0.96858 <.0001

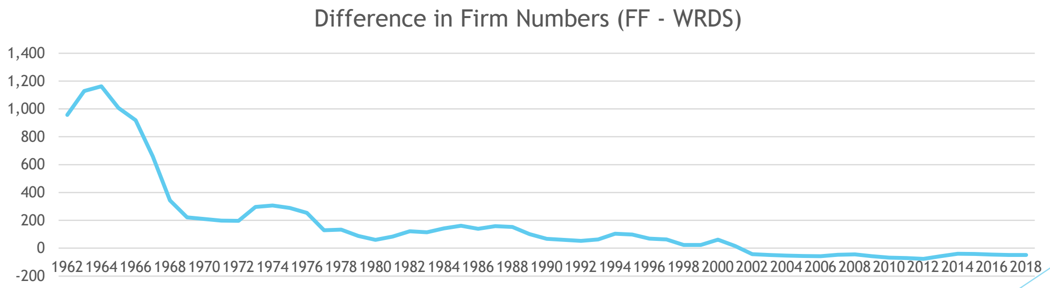
1962 - Present

Pearson Correlation Coefficients, N = 468 Prob > |r| under H0: Rho=0		
	smb	hml
WSMB	0.99280 <.0001	-0.18808 <.0001
WHML	-0.25655 <.0001	0.97669 <.0001

1980 - Present

Discrepancy or Limitations

- Though we did our best to follow the original work, there are still some 差异 discrepancy in earlier years.
  - Main reason: Fama & French replaces the missing Compustat accounting items by Moody, so that their portfolios covering more firms.



Back

CRSP Data – Rebalance (Alternative)

```
data crspm_w; set crspm; by permno date; retain me_base cumretx;
            LME=lag(meq); /*lagged month Market Values*/

            if first.permno then do;
                        LME=me/(1+retx);  me_base=LME;
                        weight_port=.; cumretx=sum(1,retx);
            end; /*1)Initiate Lagged Values*/


            if month(date)=7 then do; /*2)Portfolio Formation Date at June*/
                        weight_port=LME; /*3)RESET Weights at the beginning of July*/
                        me_base=LME;
                        cumretx=sum(1,retx);
            end;


            else if first.permno=0 then do;
                        if LME>0 then weight_port=cumretx*me_base;
                        else weight_port=.;
                        cumretx=cumretx*sum(1,retx);/*4)Dividend is not considered*/
            end;

            keep permno date retadj weight_port ME exchcd shrcd;
run;
```

- When a stock start records before July, its previous **month weights** are set to NULL
- When it is July, the previous **month weights** are set to Jun’s one and cumulative RETX is to July one
- If it is not July and not the first obs of a permno, then the market weight is calculated based on cumulative RETX and market cap on previous June.

Back

CRSP Data – Some Formulas

- Formula for Value-Weighted Returns:
  - $r_{port,t} = \sum_{j=1}^{n}\left[\left(\frac{market\ cap_{j,t-1}}{\sum_{i}^{n} market\ cap_{i,t-1}}\right) \times r_{j,t}\right]$
- Rebalance at June and dividends are not reinvested
  - “Adjusted” Shares held are not changed through July to June next year
    - $Market\ Cap_{JUL} = Market\ Cap_{JUN} \times (1 + RETX_{\ JUL})$
    - $Market\ Cap_{AUG} = Market\ Cap_{JUN} \times (1 + RETX_{\ JUL}) \times (1 + RETX_{\ AUG})$
    - …
    - $Market\ Cap_{MAY} = Market\ Cap_{JUN} \times (1 + RETX_{\ JUL}) \times (1 + RETX_{\ AUG}) \times \ldots \times (1 + RETX_{\ MAY})$
  - Note: $\frac{Market\ Cap_{AUG}}{Market\ Cap_{JUN}} = \frac{Shares_{JUN} \times PRICE_{AUG}}{Shares_{JUN} \times PRICE_{JUN}} = \frac{PRICE_{JUL}}{PRICE_{JUN}} \times \frac{PRICE_{AUG}}{PRICE_{JUL}} = (1 + RETX_{\ AUG}) \times (1 + RETX_{\ JUL})$

Slide Deck #3 Supplementary:

Event Study and Shleifer (1986) Replication

Market Efficiency and Event Study

- **Efficient Market Hypothesis** summarized in Fama (1970) provides a theoretical ground for empirical researches in Finance, Accounting, etc. to investigate the informational component in asset pricing.
  - To test this hypothesis, researchers have developed empirical tools and found interesting market phenomena regarding how market reacts to information
- **Event studies** are often used in the applications that are related to semi-strong form efficiency (information become publicly available) or strong form efficiency (private information), such as
  - Whether a given type of publicly available information might an investor use to generate higher than normal returns?
  - How much time is required for a given type of information to be reflected in asset prices?

Review: A Typical Event Study

- In a typical event study, stock return behavior is analyzed in a relative short period of time, such as **a few days** surrounding the announcement date.
- An event could be firms' specific news (e.g., M&A and Earning Announcement) or macroeconomic news (e.g., FOMC announcement)
- The goal is to measure the information impact, abnormal return, due to the event which is the difference between the stock's actual return and a proxy for the stock's return in the absence of the event.
  - a pricing model such as the CAPM or Fama-French three factor model
  - a broad market index
  - a firm characteristic-based portfolio

Horizontal Demand Curve

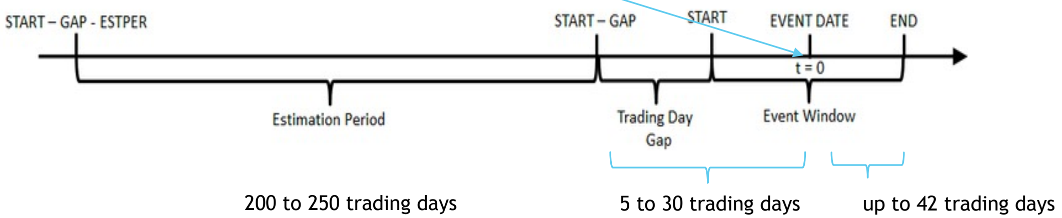
- Asset pricing models based efficient markets hypothesis (e.g., Fama and French 1993) tend to predict nearly horizontal demand curves for stocks, i.e., stock price is determined by nothing more than the factor loading *β*s
- Shleifer (1986), as well as Harris and Gurel (1986), first document a remarkable fact: when a stock is added to the index, it jumps in price by an average of 27%, and much of this jump is permanent.
- The fact that a stock jumps in value upon inclusion may indicate that the stock price can be changed through an increased demand (e.g., from index funds) even though its fundamental value does not change.

S&P Inclusion

- Constituents of the S&P 500 index can be taken out of the index because of a merger or bankruptcy and are replaced by other firms.
- Standard and Poor emphasizes that in selecting stocks for inclusion, they are simply trying to make their index representative of the US economy, not to convey any information about the level or riskiness of a firm's future cash flows
  - S&P states six criteria for inclusion: size, industry classification, capitalization, trading volume, emerging companies/industries, and responsiveness of the movements of stock price to changes in industry affairs,
  - Those criteria are all public information and none of them is concerned with the future performance of the firm.

Event Data Prep (Pricing Model)

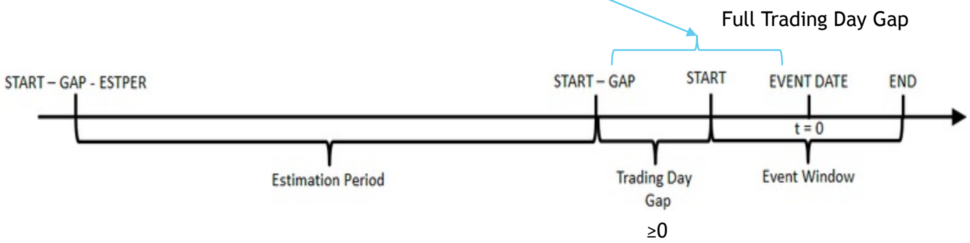
- **A precise announcement** date is required for a sample of firms
  - It is important to make sure that no other announcements are close to the ones that are investigated to avoid contaminating the estimates by confounding effect other events



- Daily returns for the sample firms and an appropriate market index are required
  - A reasonable length for entire period would be from 13 months before to a few weeks after

Event Data Prep (Pricing Model)

- A trading day gap is typically required to ascertain that the estimation period is free of any effects related to the investigated events.
  - e.g., it is well known that M&A targets experience substantial price increases during the 20 trading days preceding the actual announcement.



Trading Days vs. Calendar Days

- All the days mentioned so far are trading days, which may be tricky to mathematically imply from calendar days

2020 Trading Days Calendar

Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	Monthly Trading days	Monthly Trading hours	Quarterly Trading days	Quarterly Trading hours
Jan	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	21	136.5		
Feb	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S			19	123.5		
Mar	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	22	143.0	62	403.0
Apr	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T		21	136.5		
May	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	20	130.0		
Jun	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T		22	143.0	63	409.5
Jul	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	22	143.0		
Aug	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	21	136.5		
Sep	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W		21	136.5	64	416.0
Oct	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	22	143.0		
Nov	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M		20	127.0		
Dec	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	22	140.0	64	410.0

Jan 1	New Year's Day	Sep 7	Labor Day
Jan 20	Martin Luther King, Jr. Day	Nov 26	Thanksgiving Day
Feb 17	Washington's Birthday	Nov 27	markets close early at 1:00
Apr 10	Good Friday	Dec 24	markets close early at 1:00
May 25	Memorial Day	Dec 25	Christmas
Jul 3	Independence Day (observed)		

	Monthly Trading days	Monthly Trading hours	Quarterly Trading days	Quarterly Trading hours
total	253	1638.5	253	1638.5
avg	21	136.5	63	409.6

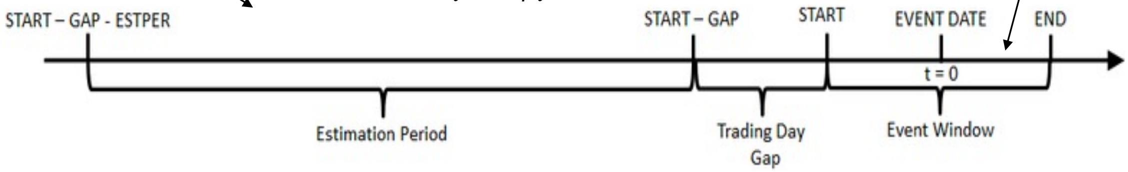
Estimation and Abnormal Returns

- The estimated coefficients of a pricing model from regressing firm-specific returns on contemporaneous market returns are used to calculate abnormal return for each firm and event.

$\hat{R}_{i,t} = \hat{\alpha}_i + \hat{\beta}_i \times R_{m,t}$

$\widehat{AR}_{i,\tau} = r_{i,t} - (\hat{\alpha}_i + \hat{\beta}_i \times R_{m,\tau})$

Estimation of $\hat{\alpha}_i$ and $\hat{\beta}_i$



Choice of Benchmarks

- Parametric Benchmarks: APT原始模型.
  - *Market Model (Not CAPM)*: $\widehat{AR}_{i,t} = R_{i,t} - (\hat{\alpha}_i + \hat{\beta}_i \times R_{m,t})$
  - *Fama French Model*: $\widehat{AR}_{i,t} = R_{i,t} - (R_{f,t} + \hat{\alpha}_i + \hat{\beta}_i(R_{m,t} - R_{f,t}) + \hat{s}_i SMB_t + \hat{h}_i HML_t)$
  - *Carhart Model*: $\widehat{AR}_{i,t} = R_{i,t} - (R_{f,t} + \hat{\alpha}_i + \hat{\beta}_i(R_{m,t} - R_{f,t}) + \hat{s}_i SMB_t + \hat{h}_i HML_t + \hat{u}_i UMD_t)$
- Non-Parametric Benchmarks:
  - *Market Adjusted Model*: Use market index as benchmark
  - *Firm-Characteristics Model*: Use matched B/M-Size(-Momentum) portfolio

Note on Benchmark Selection 首选等权投资组合

- As well documented, equal-weight portfolio rebalancing leads to an inflated return, presumable due to its heavier weights toward small stocks. On the other hand, value-weight portfolio is not always better since it is biased towards the largest market cap stocks.

- In practice, it is commonly seen that researchers use both value-weight and equal-weighted benchmark in Event Studies. However, for a series of unscheduled events occur over a long horizon, it is empirically unrealistic to assume that a preannouncement capital allocation, i.e., weighting each event by the market cap of event firms, is applicable.

- For this reason, the equal-weight portfolio returns are inherently preferred for detecting abnormal returns, though the choice of benchmark has very limited impact on event study results in short run.

❖ However, the choice of benchmark could lead to significantly different results in a long run setting. See Barber and Lyon (1997) and Mitchell and Stafford (2000) for further discussion.

The Goal: Replicating

Table I

Average Abnormal Returns Surrounding Inclusion of Stocks into the S&P 500 Index

Days relative to the Announcement Date (AD)	Average Cumulative Prediction Error: 1966–1975 (before the early warning service) N = 144	Average Cumulative Prediction Error: Sept, 1976–1983 (after the early warning service) N = 102
AD − 20 through AD − 1	−2.86	−1.49
	(−2.85)	(−1.25)
AD	−0.192	2.79
	(−0.918)	(12.4)
AD + 1 through AD + 10	−0.065	−0.859
	(−0.091)	(−1.03)
AD + 11 through AD + 20	1.12	−0.154
	(1.57)	(−0.184)

Notes:
1. *t*-statistics are included in parentheses
2. details of calculations are provided in the text

"To examine share price behavior surrounding inclusion into the S&P500, I performed a daily event study following Fama (1969) ... Specifically, the ***market model*** *is applied to describe the behavior of asset returns, using the* ***value weighted*** *market portfolio from CRSP files."*

Table II

Tests of Persistence of Abnormal Returns Subsequent to Inclusion into S&P 500

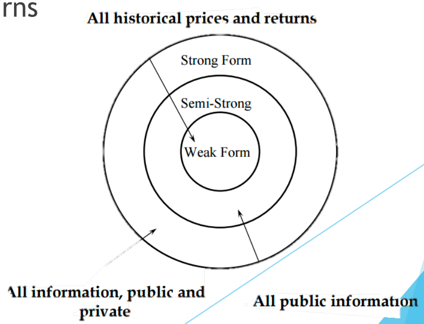
Days over which the abnormal return is cumulated	Average Cumulative Prediction Error: Full Sample Sept, 1976–1983	Early Subsample Sept, 1976–1980	Later Subsample 1981–1983
AD	2.79	2.27	3.19
	(12.4)	(7.63)	(10.1)
AD through AD + 5	2.22	1.58	2.70
	(3.46)	(1.88)	(3.10)
AD through AD + 10	1.94	1.09	2.58
	(2.20)	(0.946)	(2.17)
AD through AD + 20	1.78	−0.807	3.77
	(1.46)	(−0.505)	(2.28)
AD through AD + 60	1.71	−1.05	3.85
	(0.819)	(−0.383)	(1.36)

Notes:
1. *t*-statistics in parentheses
2. details of calculation of abnormal returns are provided in the text

Slide Deck #5: Efficient Market Hypothesis

Efficient Market Hypothesis (EMH)

- Fama (1970) : "A market in which prices always *fully reflect* available INFORMATION is called *efficient*."
- Depending the INFORMATION, there are three forms:
  - **Weak-form Efficiency:** past returns cannot predict future excess returns..
  - **Semistrong-Form Efficiency:** public information cannot be used to predict future excess returns.
  - **Strong-Form Efficiency:** no information (even direct personal knowledge of a merger, for example) can be used to predict future excess returns



Efficient Market Hypothesis (EMH)

- Maurice Kendall (1953) found no predictable pattern in stock price changes.
- Prices are as likely to go up as to go down on any particular day.
- How do we explain random stock price changes?

Efficient Market Hypothesis (EMH)

- EMH says stock prices already reflect all available information

- A forecast about favorable <u>future</u> performance leads to favorable <u>current</u> performance, as market participants rush to trade on new information.

    - Result: Prices change until expected returns are exactly commensurate with risk.

      相称的

Efficient Market Hypothesis (EMH)

- New information is unpredictable; if it could be predicted, then the prediction would be part of today's information.
- Stock prices that change in response to new (unpredictable) information also must move unpredictably.
- Stock price changes follow a <u>random walk</u>.

Random Walk: Price changes should be uncorrelated.

EMH and Competition

- Information: The most precious 商品 commodity on Wall Street
  - Strong competition 保证 assures prices reflect information.
  - Information-gathering is motivated by desire for higher investment returns.
  - The marginal return on research activity may be so small that only managers of the largest portfolios will find them worth pursuing.

Technical Analysis

- Technical analysis has been used by traders and analysts for centuries, but it has only recently achieved broad acceptance among regulators and the academic community.
- Technical analysis is a form of security analysis that uses price and volume data, which is often graphically displayed, in decision making.
- Technical analysis is used on a wide range of financial instruments, including equities, bonds, commodity futures, and currency futures.

Technical Analysis

阻碍

- An important impediment to acceptance by academics is the difficulty of capturing the subjectivity involved in technical analysis.
- The human brain can recognize, analyze, and interpret technical information that is difficult for statistical computer models to recognize and test.
- **Advantage:** Securities fraud, such as occurred at Enron Corporation and WorldCom
- **Disadvantage:** Subject to Government Intervention and Market illiquidity

Technical Analysis

- Success depends on a sluggish response of stock prices to fundamental supply-and-demand factors.
  - Supply and demand determine prices.
  - Changes in supply and demand cause changes in prices.
  - Prices can be projected with charts and other technical tools.
- Weak form efficiency
  - Relative strength
  - Resistance levels

Types of Stock Analysis

- *Fundamental Analysis* - using economic and accounting information to predict stock prices
  - Try to find firms that are *better than* everyone else's estimate.
  - Try to find poorly run firms that are not as bad as the market thinks.
  - Semi strong form efficiency and fundamental analysis

Technical vs. Fundamental Analysis

	Fundamental	Technical
Forecast Base	Facts and statistics	Price movement and charts
Horizon	Long term	Days, minutes or even shorter
Trend Analysis	No	Yes
Report	Expectation vs. Real Current vs. Historical	Trendline, Support and Resistance, etc.

When Market Efficient

- Technical Analysis Fails
  - Once a useful technical rule (or price pattern) is discovered, it ought to be invalidated when the mass of traders attempts to exploit it. In this sense, price patterns ought to be *self-destructing*.
- Fundamental Analysis Also Fails:
  - Many well-informed, well-financed firms, with similar access to similar public information, conduct such market research, and in the face of such competition it will be difficult to uncover data not also available to other analysts.

Strong Form Tests

- In legal terms, the Insiders of a firm, such as CEO, CFO and board members, possess Material and Non-public information about the firm.
- The regulators should prevent insiders from taking advantage of regular investors by trading private information, and therefore make Strong Form of EMH regulatorily impossible.
  - The SEC, for example, requires insiders to report any changes of their holding in the subject firm within two business days. The firms are called Form 3, 4, and 5.

Two reasons insider's trade

1. **Liquidity/diversification (not correlated with future outcomes)**
   - Need to fund lavish lifestyle 奢华
     - e.g., a Roman-themed 40th birthday party for your wife on the Island of Sardinia
   - Large portion of executive pay is tied up in equity
     - e.g., Les Moonves >$215 million stake in CBS as of 2/2017
2. **Private Information (systematically correlated with future outcomes)**
   - Trade-based on superior foreknowledge of company performance/future events (sell before bad news, buy before good news)

Two approaches in academic literature

1. Are the trades correlated with future outcomes?
   - **Future returns** **(Related to Efficient Market)**
   - Future operating performance (earnings); Non-performing loans; Credit rating downgrades
2. Are the trades opportunistically timed around an event?
   - Earnings Announcements; Dividend Announcement; Audit Reports; Buybacks

Opportunistic vs. Routine Insider Trading

- Cohen, Malloy, Pomorski (2012 JF)
  - A routine trader as an insider who placed a trade in the same calendar month for at least three consecutive years
  - An opportunistic traders as everyone else - those insiders for whom we cannot detect an obvious discernible pattern in the past timing of their trades.
  - Opportunistic Trading:
    - 36% of insider purchases
    - 48% of insider sales

Semi-Strong Form Tests

- Fundamental analysis uses a much wider range of information to create portfolios than does technical analysis.
- Semistrong-form market efficiency tests:
  - Whether publicly available information beyond the trading history of a security can be used to improve investment performance; as such, they are tests of.
- **Surprisingly,** several easily accessible statistics, for example, a stock's Book-to-Market or Size, seem to predict abnormal risk-adjusted returns.
  - These Findings are difficult to reconcile with the EMH and are often referred to as *efficient market anomalies*.

Speed of Information

- A fundamental principle of EMH is that any new information ought to be reflected in stock prices very rapidly.
  - When good news is made public, for example, the stock price should jump immediately.
  - A nice illustration of such rapid adjustment is provided in a study by Busse and Green (2002 JFE), who track minute-by-minute stock prices of firms that are featured on CNBC's "Morning" or "Midday Call" segments.

Post–Earnings-Announcement Price Drift

- A puzzling anomaly, therefore, is the apparently sluggish response of stock prices to firms' earnings announcements
  - Uncovered by Ball and Brown (1968 JAR)
  - Confirmed with large data by Rendleman, Jones, and Latané (1982 JFE)
- The "news content" of an earnings announcement can be evaluated by comparing the announcement of actual earnings to the value previously expected by market participants.

Weak Form Tests

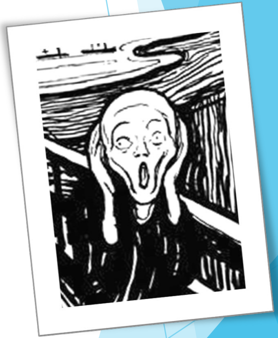
- **H0:** The information set consists only of information contained in past prices (returns).
- Campbell, Lo and MacKinlay (1997) uses value weighted and equal weighted indexes and a sample of 411 individual stock return from the CRSP database over period from 1962 to 1994.
- Autocorrelation analysis:
  - Index: small positive (first lag) return autocorrelation for index among daily, weekly and monthly horizon 自相关
  - Stock: small negative return autocorrelation for individual stocks at daily horizon

Weak Form Tests

- Returns over Short Horizon (Momentum)
  - Momentum: Good or bad recent performance continues over short to intermediate time horizons
  - Jegadeesh and Titman (1993) conclude that while the performance of individual stocks is highly unpredictable, portfolios of winners in the recent past appear to outperform
- Returns over Long Horizons (Reversal) 反转
  - Episodes of overshooting followed by correction
  - DeBondt and Thaler (1985) study found that if one were to rank the performance of stocks over a five-year: the base-period “loser” portfolio (defined as the 35 stocks with the worst investment performance) outperformed the “winner” portfolio by an average of 25% (cumulative return) in the following three-year period.

Weak Form Tests

- Violation of weak form implies violation of semi-strong and strong.
  - Question is whether the violation is large in an economic sense, stable over time, robust to different assumptions.
  - Previous studies demonstrate weak price trends, the evidence does not clearly suggest the existence of trading opportunities on individual stock level.



In Summary

- **Strong-Form Efficiency** = Illegal Trading is OK
  - Fama (1991) discusses a range of studies showing that insider trading permits generating extra profits, i.e., empirically rejecting the stron form EMH.

- **SEMI-Form Efficiency** = Fundamental Analysis does not work
  - Public information cannot be used to predict future excess returns. In an eff market, stock price should soon integrate those newly released information.
- **Weak-form Efficiency** = Technical Analysis does not work
  - Fama (1965) concluded that "so far the sophisticated analyst has escaped detection", after having studied the performance of 39 mutual funds from 1950 t 1960.

Major Takeaways

- Failure of the EMH in empirical tests may be caused by one or combination of following:
  - A failure of informational efficiency
    - In its rational expectation setting, the forecast errors, that is, $\varepsilon_{t+1} = P_{t+1} - E_t P_{t+1}$ should be zero **on average** and should be uncorrelated with **any information** that was available at the time the forecast was made.
  - An inappropriate choice of the model for equilibrium returns, or
    - an economic model is needed to determine the required returns (basec on $E_t P_{t+1}$)
  - simply the EMH does not hold in the 'real world'.

Fund Fees

- An expense ratio is the amount companies charge investors to manage a mutual fund or exchange-traded fund (ETF).

$$Expense\ Ratio = \frac{Operating\ Expenses}{Avg.AUM}$$

  - According to Morningstar, the average ETF expense ratio in 2016 was 0.23%, compared with the average expense ratio of 0.73% for index mutual funds and 1.45% for actively managed mutual funds.
- Hedge funds operate very differently, with both a management fee and performance fee.
  - According to Financial times, the average hedge fund earns 1.67% in management fees and is paid 18% investment profits in 2016. Investors pays 50% of total return and 80% of alpha to hedge funds from 2007 to 2016.

If Efficient Market Hypothesis is True

- A great deal of the activity of these fund managers—the search for undervalued securities—is at best wasted effort, and quite probably harmful to clients because it costs money and leads to imperfectly diversified portfolios.
- Investor should instead stay with passive strategies that aim only at establishing a well-diversified portfolio of securities without attempting to find under- or overvalued stocks.
- Consequently, the EMH has never been widely accepted on Wall Street

Mutual Fund Performance

- The conventional performance benchmark today is a four-factor model, which employs:
  - the three Fama-French factors (the return on the market index, and returns to portfolios based on size and book-to-market ratio)
  - plus a momentum factor (a portfolio constructed based on prior-year stock return).

Market Efficiency Issues

- Selection Bias Issue
  - Only unsuccessful investment schemes are made public; good schemes remain private.
- Lucky Event Issue
  - If many investors using a variety of schemes make fair bets, statistically speaking, some of those investors will be lucky and win a great majority of the bets
    - the successful (lucky) investor will call it skill

Mutual Fund Performance Persistence

- Carhart (1997 JoF) – there is only minor persistence in relative performance across managers. Moreover, much of that persistence seems due to expenses and transactions costs rather than gross investment returns.
- Bollen and Busse (2004 RFS) – support for performance persistence over short time horizons
- Berk and Green (2004 JPE) – skilled managers will attract new funds until the costs of managing those extra funds drive alphas down to zero.

Slide Deck #6:

Event Study

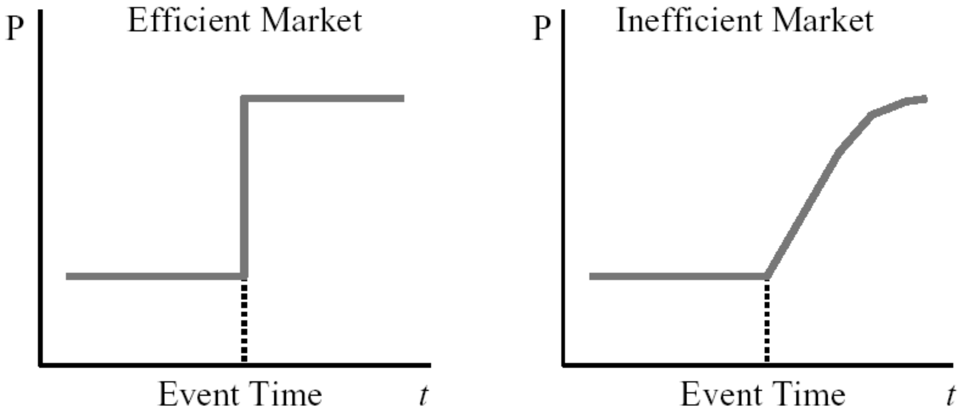
Who needs Event Study

- Event studies examine the behavior of firms' stock prices around corporate events
    - Its importance and wide usage are evident from the academic literature and consulting reports.
    - It includes events as earnings announcements; releases of quarterly and annual financial reports; announcements of dividends, stock splits, mergers, etc.
- Mastering the ESA is essential to researchers and practitioners.

Efficient Market Hypothesis

- According to **Semi-Strong Form Efficient Market Hypothesis**
  - Security prices should quickly reflect all currently available information,
  - Price changes can, therefore, be used to measure the importance of an event of interest
- The notion of EMH leads to a powerful research methodology: **Event Study**
  - Empirical financial research enables us to assess the impact of a particular event on a firm's stock price.

Efficient Market Hypothesis



Event Study Background

- **Event studies** can play an important an important role not only for EMH
- Measure the impact of certain event to the <u>equity value of a firm</u> and the <u>wealth of the firms' claimholders</u> (if market is rational)
- The event that affects a firm's valuation may be
  - within the firm's control, e.g. M&A, SEO, stock split
  - Outside the firm's control, such as a macroeconomic announcement

Event Study Background

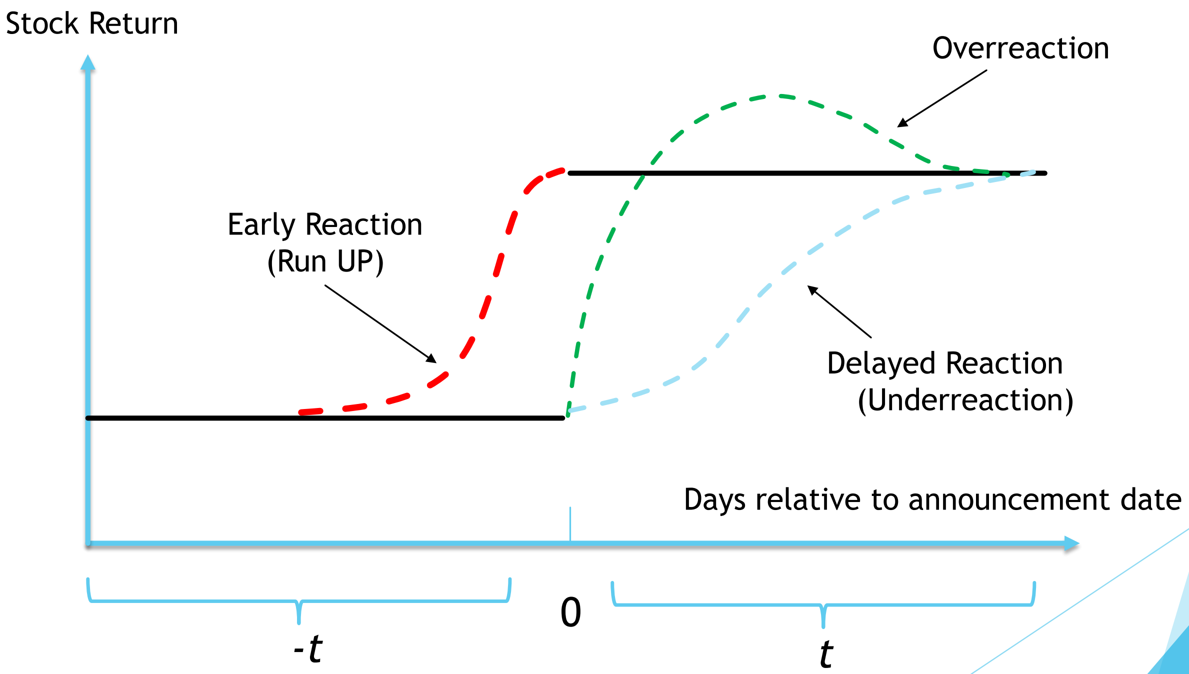
起诉

- Extensive use in litigation and policy discussions
  - Detect leaked information by pre-announcement nonzero abnormal return
  - Used as evidence of irrationality
- Many famous results in finance derived from event studies

Short Term vs. Long Term Event Study

- So far we have showed “Short-Run Term” Event Study
  - Time Spans: Days (usually <= 90 calendar days)
- As you can imply, there are Long-Run Event Study
  - Very Sensitive to Market Models due to long-term Compounding effect
  - Overlap and clustering of windows becomes important
    - Solution: Calendar Time Event Study
  - Statistical properties of the estimators becomes important

Market Efficiency: Reaction to + news



Not A trivial Task

- On any day, stock prices respond to a wide range of economic news such as updated forecasts for GDP, inflation rates, interest rates, or corporate profitability.
- Isolating the part of a stock price movement that is attributable to a specific event is not a trivial exercise.
- The general approach starts with a proxy for what the stock's return would have been in the absence of the event.

How to Measure Norm

- The abnormal return due to the event is the difference between the stock's actual return and a proxy for the stock's return in the absence of the event.
- The stock's return in the absence of the event can be measured by a benchmark
- Several methodologies for estimating the benchmark return:
  - a broad market index
  - an asset pricing model such as the CAPM or Fama-French three factor model

Benchmark: Market Index

- Definition: Stock *i*'s Abnormal Return, $AR_{i,t}$, is defined as the deviation of its return from *its Normal Return* $NR_{i,t}$.

$$\widehat{AR}_{i,t} = R_{i,t} - \widehat{NR}_{i,t}$$

- <u>*Market Adjusted Model*</u>: Use market index as benchmark
- One basic of benchmark is a market benchmark, such as the Standard & Poor's 500 Stock Price Index (S&P 500) or the Center for Research in Security Prices (CRSP) composite value-weighted or equal-weighted indices in the United States.

Benchmark: Factor Models

- Market Model (Not CAPM):

$$\widehat{NR}_{i,t} = \hat{\alpha}_i + \hat{\beta}_i \times R_{m,t}$$

- Accordingly, $ar_{i,t}$ is estimated by plugging this $\widehat{nr}_{i,t}$ into equation

$$\widehat{AR}_{i,t} = r_{i,t} - \left(\hat{\alpha}_i + \hat{\beta}_i \times R_{m,t}\right)$$

where $\widehat{ar}_{i,t}$ is actually the error term, $\varepsilon_{i,t}$, of the following regression:

$$R_{i,t} = \hat{\alpha}_i + \hat{\beta}_i \times R_{m,t} + \varepsilon_{i,t}$$

*A method established by Fama, Fisher, Jensen and Roll (1969)

Benchmark: Factor Models

- Suppose that the analyst has estimated that $\hat{\alpha}_i = 0.05\%$ and $\hat{\beta} = 0.8$
- On a day that the market goes up by 1%, you would predict from market model that the stock should rise by an expected value of
$$0.05\% + 0.8 \times 1\% = .85\%$$
- If the stock actually rises by 2%, the analyst would infer that firm-specific news that day caused an additional stock return of
$$\widehat{AR}_{i,t} = 2\% - .85\% = 1.15\%$$

The Joint-Test Problem for Pricing Models

- The Event Study is a set of tools to measure the <u>magnitude and significance of abnormal security price (test 1)</u>.
- However, to measure abnormal, we need to obtain the norm of security price, expected returns, by certain <u>asset pricing model(test 2).</u>
  - Market Model (CAPM) or FF3 is assumed to be an appropriate model

Benchmark: Factor Models

- Fama French Model:

daily

$$\widehat{AR}_{i,t} = R_{i,t} - \left(R_{f,t} + \hat{\alpha}_i + \hat{\beta}_i\left(R_{m,t} - R_{f,t}\right) + \hat{s}_i SMB_t + \hat{h}_i HML_t\right)$$

- Fama French + UMD Model:

$$\widehat{AR}_{i,t} = R_{i,t} - \left(R_{f,t} + \hat{\alpha}_i + \hat{\beta}_i\left(R_{m,t} - R_{f,t}\right) + \hat{s}_i SMB_t + \hat{h}_i HML_t + \hat{u}_i UMD_t\right)$$

Cumulative Abnormal Returns

- Definition: Stock i's CAR for a period is the sum of its ARs over that period.

- CAR from period s to t is estimated as following:

$$\widehat{CAR}_{i,s,t} = \sum_{\tau=s}^{t} \widehat{AR}_{i,\tau}$$

where $\widehat{CAR}_{i,s,t}$ is stock i's estimated CAR from period s to t, and $\widehat{AR}_{i,\tau}$ is stock i's estimated AR at period τ.

Abnormal Conversion

- Calculate Abnormal Return

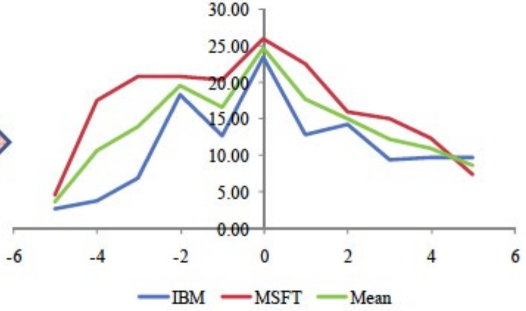
Simply market adjusted

Stock 1: IBM			Stock 2: MSFT		
Date	Event day	Return (%)	Date	Event day	Return (%)
12/26/2008	-5	4.95	8/20/2008	-5	-4.63
12/29/2008	-4	1.13	8/23/2008	-4	4.78
12/30/2008	-3	1.38	8/24/2008	-3	-0.39
12/31/2008	-2	2.34	8/25/2008	-2	2.93
1/1/2009	-1	1.92	8/26/2008	-1	3.82
1/2/2009	0	7.79	8/27/2008	0	3.72
1/5/2009	1	-1.75	8/30/2008	1	0.03
1/6/2009	2	2.24	8/31/2008	2	1.31
1/7/2009	3	2.40	9/1/2008	3	3.21
1/8/2009	4	-1.48	9/2/2008	4	4.42
1/9/2009	5	-0.88	9/3/2008	5	-2.16

Stock 1: IBM			Stock 2: MSFT		
Date	Event day	Return (%)	Date	Event day	Return (%)
12/26/2008	-5	2.67	8/20/2008	-5	4.60
12/29/2008	-4	1.11	8/23/2008	-4	13.01
12/30/2008	-3	3.18	8/24/2008	-3	3.26
12/31/2008	-2	11.29	8/25/2008	-2	-0.01
1/1/2009	-1	-5.55	8/26/2008	-1	-0.41
1/2/2009	0	10.77	8/27/2008	0	5.55
1/5/2009	1	-10.58	8/30/2008	1	-3.43
1/6/2009	2	1.39	8/31/2008	2	-6.62
1/7/2009	3	-4.92	9/1/2008	3	-0.82
1/8/2009	4	0.31	9/2/2008	4	-2.79
1/9/2009	5	0.09	9/3/2008	5	-4.86

Some Graphic Presentation

Stock 1: IBM			Stock 2: MSFT		
Date	Event day	Return (%)	Date	Event day	Return (%)
12/26/2008	-5	2.67	8/20/2008	-5	4.60
12/29/2008	-4	1.11	8/23/2008	-4	13.01
12/30/2008	-3	3.18	8/24/2008	-3	3.26
12/31/2008	-2	11.29	8/25/2008	-2	-0.01
1/1/2009	-1	-5.55	8/26/2008	-1	-0.41
1/2/2009	0	10.77	8/27/2008	0	5.55
1/5/2009	1	-10.58	8/30/2008	1	-3.43
1/6/2009	2	1.39	8/31/2008	2	-6.62
1/7/2009	3	-4.92	9/1/2008	3	-0.82
1/8/2009	4	0.31	9/2/2008	4	-2.79
1/9/2009	5	0.09	9/3/2008	5	-4.86



Focus on Specific Event Windows

Stock 1: IBM			Stock 2: MSFT		
Date	Event day	Return (%)	Date	Event day	Return (%)
12/26/2008	-5	2.67	8/20/2008	-5	4.60
12/29/2008	-4	1.11	8/23/2008	-4	13.01
12/30/2008	-3	3.18	8/24/2008	-3	3.26
12/31/2008	-2	11.29	8/25/2008	-2	-0.01
1/1/2009	-1	-5.55	8/26/2008	-1	-0.41
1/2/2009	0	10.77	8/27/2008	0	5.55
1/5/2009	1	-10.58	8/30/2008	1	-3.43
1/6/2009	2	1.39	8/31/2008	2	-6.62
1/7/2009	3	-4.92	9/1/2008	3	-0.82
1/8/2009	4	0.31	9/2/2008	4	-2.79
1/9/2009	5	0.09	9/3/2008	5	-4.86

Add up the abnormal returns

	Stock 1: IBM	Stock 2: MSFT
CAR [-1,0]	5.22	5.14

Long Run Event Studies

- Conventionally, an event analysis is considered as a long-horizon event study when it involves an event window of one year or more.
- Problems: The systematic errors that arise with imperfect expected return proxies are compounded over long horizons.
  - Risk adjustment based on historical estimates could be significantly biased
  - Estimates are highly sensitive to model choice
- Two approaches have commonly been employed to measure long-run abnormal stock returns after corporate events:
  - Buy-and-hold abnormal return (BHAR) method Portfolio constructions
    - true magnitude of returns of an investment strategy
  - Calendar time portfolio method (Jensen's alpha). FF3 - model
    - controls well for cross-sectional dependence among sample (more reliable inferences)

BHAR Approach in General

- Two common ways to measure abnormal returns over a given period:
  - Cumulative Abnormal Return (CAR): Short-Run

$$CAR_{it} = \sum_{t=1}^{T} (R_{it} - E(R_{it}))$$

  - Buy-and-Hold Abnormal Return (BHAR): Long-Run

$$BHAR_{it} = \prod_{t=1}^{T} (1 + R_{it}) - \prod_{t=1}^{T} (1 + E(R_{it}))$$

- In long run, the difference between CAR and BHAR is no longer negligible
  - Barber and Lyon (1997) demonstrate that the BHAR tends to be slightly lower than CAR, but becomes dramatically greater than the annual CAR when annual BHAR increases beyond 28%.
  - CAR is a sum of monthly (daily) abnormal returns (AR), a test of the null hypothesis, (mean CAR is equal to 0), is not an ideal test of the null hypothesis for a long-term event study

Choice of Long Run Benchmark

- A standard way to create a matching portfolio, such as Barber and Lyon (1997), is the following:
  - From each July to the following June, all CRSP common stocks are split into size decile portfolios according to the market capitalization of common stocks listed in NYSE at the previous June.
  - Within each size decile, firms are sorted into quintiles on the basis of their book-to-market ratios prior to or in the December in the year before. This lag treatment allows for delays in the reporting of financial statements.
  - The benchmark portfolios should exclude event firms, but they include all securities that can be assigned to a characteristic-based matched portfolio.

Benchmark Rebalance

- There are two ways to calculate long-run return of the size and book-to-market reference portfolios:
  - Rebalancing Portfolio: $\prod_{t=1}^{T}(1+E(R_{it})) = \prod_{t}^{T}\left(1+\sum_{i=s}^{n_t} w_{it}R_{it}\right)$
  - Fixed Portfolio: $\prod_{t=s}^{T}(1+E(R_{it})) = \sum_{i=s}^{n_t}(w_{is}\prod_{t=1}^{T}(1+R_{it}))$
- Mitchell and Stafford (2000) argues fixed portfolio approach will not involve any newly listed security subsequent to period s, or any monthly rebalancing issues of the portfolio.
  - Equal weighted rebalance tends to overstay
  - Value weighted rebalance tends to buy (sell) more past winner (loser)
- Note: For a series of unscheduled events, it is empirically unrealistic to assume that a preannouncement capital allocation, i.e., weighting each event by the market cap of event firms, is applicable. Therefore, the equal-weight portfolio returns are inherently preferred for detecting abnormal returns

Calendar Time Portfolio Approach

- An event portfolio is formed each period to include all companies that have completed the event within the prior *n* periods (such as 3 o 5 years).
  - An equal-weighted (or value-weighted) portfolio is constructed from all sample firms that participated in the event within the previous *n* years.
  - The portfolio is rebalanced monthly to drop all companies that reach the end of their *n*-year period and add all companies that have just experience their events.

$$R_{i,t} - R_{f,t} = \hat{\alpha}_i + \hat{\beta}_i\left(R_{m,t} - R_{f,t}\right) + \hat{s}_i SMB_t + \hat{h}_i HML_t$$

  - Inferences about the abnormal performance are on the basis of the estimated $\hat{\alpha}_i$ and its statistical significance. Since $\hat{\alpha}_i$ is the average monthly abnormal performance over the T-month post-event period, it can be used to calculate annualized post-event abnormal performance.