Course Outlines

Behavioral Finance

Summer Semester 2002

University of Lüneburg

Frank A. Schmid

mail@frankschmid.com

http://frankschmid.com

March 25, 2002

Course Content

Introduction

1. Noise
2. Expected Utility and Risk Aversion
3. Heuristics and Biases
4. Choice, Values, and Frames
5. The Sharpe-Lintner CAPM
6. Efficient Markets
7. Predictability of Returns
8. Judging Investment Strategies

Empirical Studies

9. Excessive Volatility
10. The Closed-End Fund Puzzle *
11. Stock Market Overreaction (I)
12. Stock Market Overreaction (II) *
13. Long-Run Performance of IPOs
14. Stock Market Underreaction (Drift) *
15. Predictability of Corporate Earnings

Theoretical Models

16. Noise Trader Risk (Limits of Arbitrage I)
17. Professional Arbitrage (Limits of Arbitrage II) *
18. Feedback Strategies (Bubbles) *

Case Study

19. The Valuation of 3Com and Palm *

Growth-Optimal Investing

20. Stock Market Valuation
21. Random Walk versus Mean Reversion

* Time permitting

Literature

[1] Shleifer, Andrei (2000) *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford (UK): Oxford University Press.
[2] Taleb, Nassim Nicholas (2001) *Fooled by Randomness: The Hidden Role of Chance in the Markets and in Life*. New York: Texere.
[3] Thaler, Richard H., ed. (1993) *Advances in Behavioral Finance*. New York: Russell Sage Foundation.
[4] Tversky, Amos, and Daniel Kahneman (1974) “Judgement under Uncertainty: Heuristics and Biases,” *Science* 185, 1124-31, reprinted in: Daniel Kahneman, Paul Slovic, and Amos Tversky (1982), eds., *Judgement under Uncertainty: Heuristics and Biases*. Cambridge (UK): Cambridge University Press, 3-20.
[5] Kahneman, Daniel, and Amos Tversky (1984) “Choices, Values, and Frames,” *American Psychologist* 39, 341-50, reprinted in: Daniel Kahneman and Amos Tversky (2000), eds., *Choices, Values, and Frames*. Cambridge (UK): Cambridge University Press, 1-16.

Introduction

1. Noise

Reference:

Black, Fischer (1986) “Noise,” *Journal of Finance* 41, 529-43.

Knight, Frank H. (1921) “The Meaning of Risk and Uncertainty,” Part III, Chapter VII from *Risk, Uncertainty, and Profit*. Boston: Houghton and Mifflin. Online edition: http://www.econlib.org/library/Knight/knRUP.html.

Taleb, Nassim Nicholas (2001) *Fooled by Randomness: The Hidden Role of Chance in the Markets and in Life*. New York: Texere.

But wise men perceive approaching things

> Because gods perceive future things, men what is happening now, but wise men perceive approaching things.
>
> Philostratus, Life of Apollonius of Tyana, VIII, 7.

“Men know what is happening now.
The gods know the things of the future,
the full and sole possessors of all lights.
Of the future things, wise men perceive
approaching things. Their hearing
is sometimes, during serious studies,
disturbed. The mystical clamor
of approaching events reaches them.
And they heed it with reverence. While outside
on the street, the people hear nothing at all.”

--Constantine P. Cavafy (1915), Greek poet (1863-1933)

Noise versus information (noise versus signal)

Table P.1 Table of Confusion

Presenting the central distinctions used in the book

GENERAL	
Luck	Skills
Randomness	Determinism
Probability	Certainty
Belief, conjecture	Knowledge, certitude
Theory	Reality
Anecdote, coincidence	Causality, law
Forecast	Prophecy
MARKET PERFORMANCE	
Lucky idiot	Skilled investor
Survivorship bias	Market outperformance
FINANCE	
Volatility	Return (or drift)
Stochastic variable	Deterministic variable
PHYSICS AND ENGINEERING	
Noise	Signal
LITERARY CRITICISM	
None (literary critics do not seem to have a name for things they do not understand)	Symbol
PHILOSOPHY OF SCIENCE	
Epistemic probability	Physical probability
Induction	Deduction
Synthetic proposition	Analytic proposition

Source: Taleb (2001, p. 3).

Noise is what makes our observations about the real world imperfect

When trying to draw inferences from observations, individuals are faced with a signal extraction problem

The lower the signal-to-noise ratio, the harder it is to extract information from observations

The process of extracting information from noise is called filtering (see figures below).

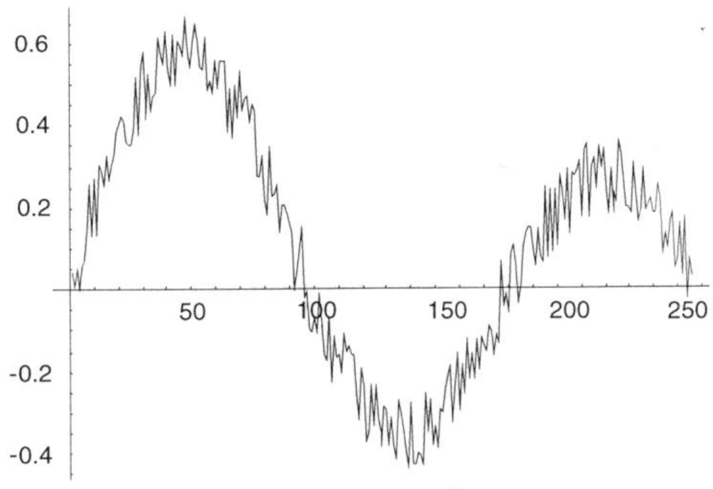


Figure 11.1 Unfiltered Data Containing Signal and Noise.

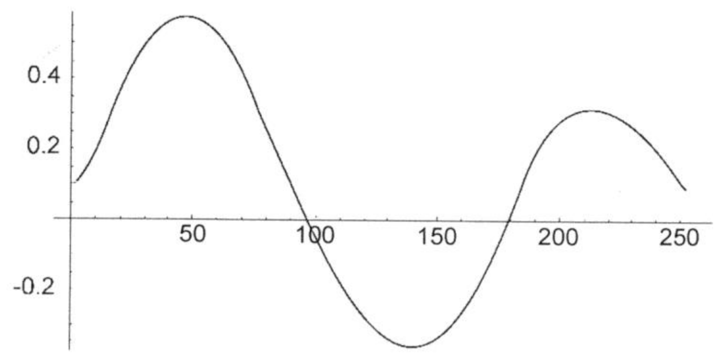


Figure 11.2 Same Data With its Noise Removed.

Source: Taleb (2001, p. 167).

Individuals frequently make decisions based on noise rather than information (signal)

Individuals who cannot distinguish between noise and information are called <u>noise traders</u>

Individuals who are able to distinguish signal from noise, we call <u>informed investors</u>

Informed investors trade less frequently than noise traders, because information is a <u>strict subset</u> of the union of noise and information.

Noise in asset prices

When we think of a financial asset, we think of the log prices following a continuous sample path in time, $p(t)$

The stochastic process $p(t)$ can be described by

$$p(t) = \mu \cdot t + \sigma \cdot B(t)$$

where $B(t)$ is standard Brownian motion, and μ, σ describe the (mean and standard deviation of the) normal distribution of the log returns of the asset in question

In Monte Carlo studies we can trace out sample paths of asset prices, called histories

From a high number of (an infinite number of possible) histories, we can draw statistical inferences about the wealth a given asset (portfolio) generates for an investor over time

Assume a financial asset with an expected excess (i.e., in excess of the risk-free rate of return) return (μ) of 15 percent and a standard deviation (σ) of 10 percent

Given these risk-return characteristics, we expect close to 68 out of 100 sample paths to fall within a band of plus and minus 10 percent around the 15 percent excess return

A 15 percent return per annum with a 10 percent standard deviation translates into a 93 percent probability of making money in any given year (probability of "being up" for the year)

However, seen on a narrow time scale, this translates into a mere 50.02 percent probability of making money over any given second (see the following table).

Table 3.1 Probability of making money at different scales.

Scale	Probability
1 year	93%
1 quarter	77%
1 month	67%
1 day	54%
1 hour	51.3%
1 minute	50.17%
1 second	50.02%

Source: Taleb (2001, p. 57).

Thus, watching a portfolio over a narrow time increment reveals next to nothing.

Human decision-making is plagued with errors, some of which are listed below, and some more we address in the subsequent chapters "Heuristics and Biases (Descriptive Analysis of Decision-Making I)" and "Choice, Values and Frames (Descriptive Analysis of Decision-Making II)"

- The with-without principle
- Assume that the Federal Reserve keeps cutting interest rates in response to a slowing economy, yet, the economy keeps slowing
    - "The interest rate cuts prove ineffective" is a frequent journalistic comment in times of monetary easing in the face of economic slowing
        - A correct analysis would have to compare the economic situation in question with one in which the Federal Reserve did not cut interest rates (and in which the economy possibly would have slowed even more).

Related to the with-without problem is the problem of hidden success

Success often times does not go on record because it is hidden

The Bank of Japan of giving cause (or at least, exacerbating) to the asset price bubble that Japan experienced in the late 1980s

The Bank of Japan defended itself as follows (Yutaka Yamaguchi, 1999, *Asset Price and Monetary Policy: Japan's Experience*, Symposium "New Challenges for Monetary Policy", Federal Reserve Bank of Kansas City, http://www.kc.frb.org/PUBLICAT/SYMPOS/1999/S99yama.htm):

> "Japan's inaction was not without reasons, as inflation was almost nil, but was read as expressing our emphasis on international cooperation. A bizarre argument gained strength that Japan, as the largest creditor, maintain low rate as the cornerstone for global stability and growth. Easy money was also regarded necessary to achieve a national economic goal of the day: reduction of external surplus through growth of domestic demand.
>
> "For all of these reasons, policy-makers felt that, for their actions to gain public support, they would need absolutely solid evidence of inflationary potential building up. When such evidence finally emerged in 1989, they quickly seized the moment to reverse the policy. Inflation picked up but peaked in 1990-01 at mid-3 percent. On the basis of the inflation track record, I would conclude that pre-emptive policy to achieve sustained price stability would probably have been desirable but must have been very hard to initiate. Professors Bernanke and Gertler indicated in their simulation results that the Bank of Japan should have raised short-term interest

> rates to 8 percent in 1988 and even higher in the following years. But all through 1988 inflation was pretty close to zero. I don't see how a central bank can increase interest to 8 or 10 percent when we don't have inflation at all."

Had the Bank of Japan hiked the short-term interest as drastically as suggested by Bernanke and Gertler, it would for sure have attracted harsh criticism from the business community and politicians for strangling an otherwise healthy economy

> Sadly enough, assume that Bernanke and Gertler are right in that the Bank of Japan should have hiked interest rates drastically at a time where real growth was high and stable and the rate of inflation was low and stable. The <u>success</u> the Bank would have achieved in preventing the stock market bubble, which popped in early 1990 and has put a drag on Japan's economic growth to this day, <u>would have never been recognized</u>.

False implications

From logic we know:

$$(A \Rightarrow B) \Leftrightarrow (\overline{A} \Leftarrow \overline{B})$$

Note that $(A \Rightarrow B)$ tells us nothing about whether $(B \Rightarrow A)$ is empirically true

For instance, the statement

"If conditions are right, I will close on the contract"

contains no information on what the individual will do if conditions are not right.

Type I and type II errors

In judging the information content of predictions, it not only matters how often someone is right, it also matters how often someone is wrong

After stock market crashes newspapers are full of reports about self-proclaimed gurus who "predicted" (prophesized) the crash

If someone keeps predicting that a stock market crash is imminent, he will eventually be right (presumed he lives long enough)

Note that even a broken clock tells the correct time twice a day.

Survivorship bias

We heard of Alexander the Great and Julius Cesar because they took considerable risk and ended up succeeding in their endeavors

How about other people in history who were equally intelligent and courageous but happened to be unlucky?

Unfortunately, the history of the unlucky remains unwritten.

Heroes are heroes because they are heroic in behavior, not because they won or lost

It is the generator that matters, not the result

For instance, people tend to call investors successful because they make money (result), not because they have a superior investment strategy (generator), thus ignoring the role of chance and the survivorship bias

Many of those that have an equally foolish investment style have gone broke, which introduces a bias when the investment strategy's performance is measured by the performance of the survivors.

Causality or common cause?

Taleb (2001, p. 53) nicely illustrates the problem when writing …

"On the rare occasion when I boarded the 6:42 [a.m.] train to New York I observed with amazement the hordes of depressed business commuters (who seemed to have preferred to be elsewhere) studiously buried in the *Wall Street Journal* … Indeed it is difficult to ascertain whether they seem depressed because they are reading the newspaper, or if depressive people tend to read the newspaper, or if people who are living outside their genetic habitat both read the newspaper and look sleepy and depressed"

The problem "causality or common cause?" is particularly thorny when one observation precedes the other

- It is widely accepted that periods of high rates of growth of monetary aggregates (e.g., M2, MZM) precede periods of high rates of inflation
  - The regularity at which inflation follows money growth led Nobel laureate Milton Friedman conclude that "Inflation is always and everywhere a monetary phenomenon"
    - Fisher Black (1987) argues that individuals might respond to higher rates of <u>expected</u> inflation with higher money demand
      - For the naïve observer, it looks as if money growth caused inflation, although money growth and inflation have a common, hidden cause.

Observational equivalence

Two (or more) possibly very different processes might generate identical data

- For instance, both collusion (in an oligopolistic market) and perfect competition generate the same price observations, both from a static and from a dynamic perspective
  - The fact that gas prices are identical (and, consequently, move simultaneously) across gas stations in a given neighborhood conforms both with perfect competition and with collusion
    - Ironically, if gas prices are not identical or do not move simultaneously then there is indeed reason to believe that perfect competition does not apply, yet, journalists often take this as a sign that competition works.

Be conservative!

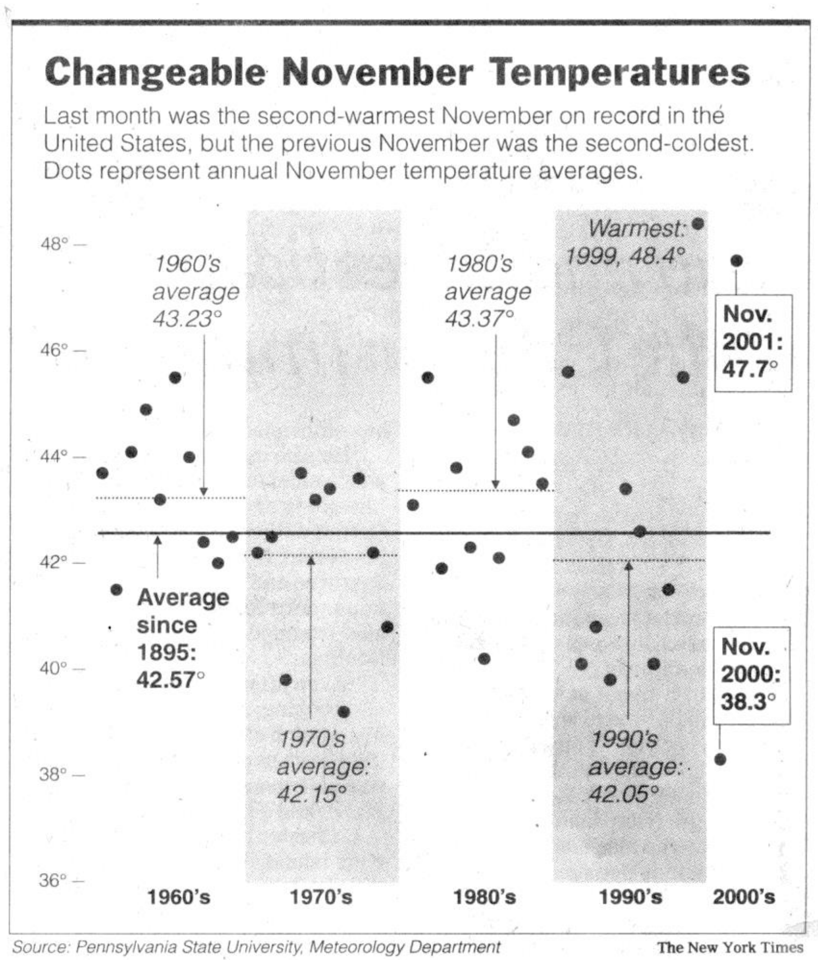
Which of the following is a random process more likely to generate?

+,+,+,+,+,+

+,-,+,-,+,-,+

Clearly, a random process generates the top series with the same probability as the bottom series

Do a couple of hot summers give proof of global warming?

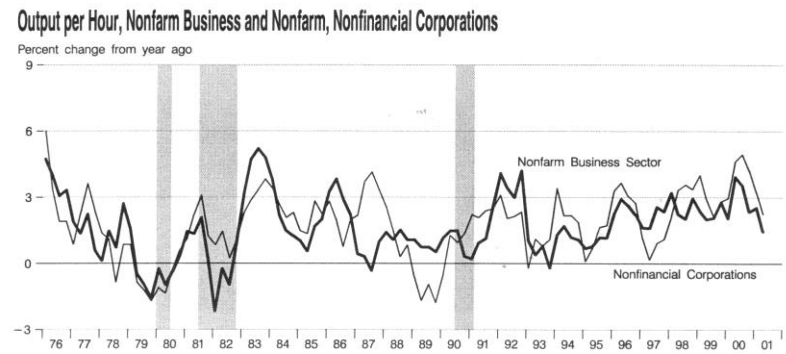


Source: *The New York Times*, Saturday, December 8, 2001.

Do a couple of positive earnings surprises warrant a rating upgrade, e.g., to "strong buy" from "buy"?

See the chapter "Predictability of Corporate Earnings."

Do a couple of years of above-average productivity growth rates tell about a “new economy”?



Source: National Economic Trends, Federal Reserve Bank of St. Louis, November 2001.

Note that the geometric mean in annual growth rates of U.S. labor productivity over the period 1947-2000 runs at about 2.2 percent.

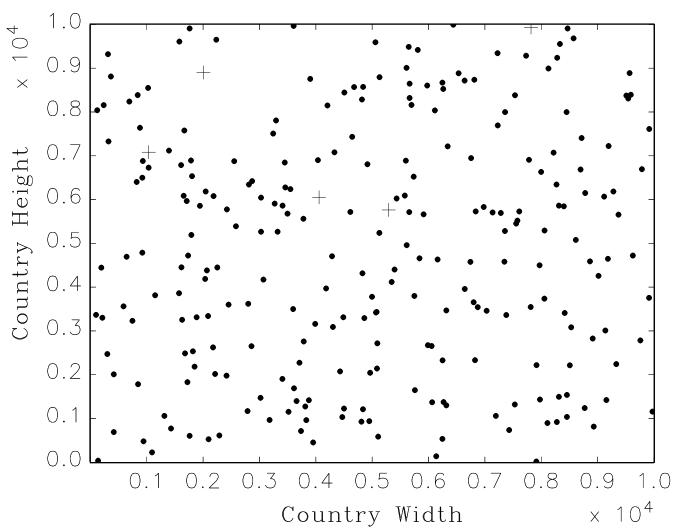
Also, <u>random series rarely look random</u>

The following graph shows a square country that is divided into 100 million (10,000 by 10,000) squares, all of which are inhabited by one citizen

I randomly chose five squares for nuclear power plant location (drawn without replacement; plus signs)

I also randomly chose 250 of the country's citizens to fall ill from leukemia (drawn without replacement; solid circles)

Note the (random) clusters of cancer patients



Selection bias

On January 2nd you get a letter from a manager of an offshore fund in which he gives testimony of his market timing abilities by informing you that the market will go up during the month; the prediction turns out to be true

Then you receive another letter on February 1st telling you that the market will go down, which again proves to be true

Then you get another letter on March 1st, …

After a couple of months, intrigued by the accuracy of the predictions, you are ready to invest into the fund

> You put all your savings in the fund, just to see your money melt away like snow in the sun.

What happened?

> The con operator started out sending 10,000 letters to addresses he took out of the phone book; 5,000 letters predicted a market decline, 5,000 predicted a rise
>
> Those (and only those) who had received the correct prediction for January, received a letter with a prediction for February; again half the letters predicted a decline, half of them predicted a rise; and so on.

Remain skeptical!

> "A man may act upon an estimate of the chance that his estimate of the chance of an event is a correct estimate. To be sure, after the decision is made he will be likely to sum all up in a certain degree of confidence that a certain outcome will be realized, and in practice may go farther and assume that the outcome itself is a certainty."
>
> Frank Knight (1921, III.VII.41)

Knight points out that our judgement on the likelihood of events is prone to error, and might change after the decision-making, which can lead to <u>probability blindness</u>

Related to Frank Knight's skepticism about judgement calls is Karl Popper's skepticism about scientific evidence

According to Karl Popper, there are two types of theories:

- Theories that are known to be wrong, as they were tested and adequately rejected (falsified).
- Theories that have not yet been known to be wrong, not falsified yet, but are exposed to be proved wrong.

In summary, investors should be skeptical about quantitative models of risk management (such as Value-at-Risk, VAR)

There is the risk that is captured by the model, but there is also the risk that this model is inadequate (and doomed to be falsified by the data).

Bayesian inference

Bayesian inference is not always intuitive, which leads to poor updating of beliefs to new information

Taleb (2001, 154-156 “Kafka in a Courtroom”) reports that during the O.J. Simpson trial, one of the defense lawyers (and Harvard professor!) presented the argument that only 10 percent of men who brutalize their wives (which O.J. did for a fact) go on to murder them

Bayesian inference suggests that this statement is a grossly inadequate perspective

In fact, the probability that O.J. is the murder equals

$$\pi_{s \cdot m} = \frac{j_{sm}}{q_m} = \frac{0.1}{0.2} = 0.5$$

where …

s is the state that the woman was murdered by her husband;

m is the message that the murdered wife was previously brutalized by her husband (but not necessarily murdered by him);

$\pi_{s \cdot m}$ is the probability that the husband is the murderer given the information that he previously brutalized her;

j_{sm} ($=10$ percent) is the joint probability of murder and physical marital abuse mentioned by the lawyer;

q_m ($=20$ percent) is the probability that murdered women were brutalized by their husbands (but not necessarily killed by them).

Another, must less intuitive example is the Monty Hall Problem, which is named after the TV game show "Let's Make a Deal," that used to be hosted by Monty Hall but has been taken off the air; see http://math.rice.edu/~ddonovan/montyurl.html

The contestant is given the opportunity to select one closed door of three, behind one of which there is a prize; the other two doors hide goats ("no prize")

Once the contestant has made her selection, Monty Hall opens one of the remaining doors, revealing that it does not contain the prize

Monty Hall then asks the contestant if she would like to switch doors

Although it is not intuitive (to most people), the contestant should switch!

There are two ways to go about the problem (assuming that the contestant chose door #1 and Monty Hall opened door #2)

First, remember that the unconditional probability that the prize is behind door #1 equals 1/3; given that the prize is not behind door #2, the probability that the prize is behind door #3 equals 2/3

Second—and more elegantly—apply Bayesian inference to update your belief

$$\pi_{s \cdot m} = \frac{j_{sm}}{q_m} = \pi_s \frac{q_{m \cdot s}}{q_m}$$

where π_s is the prior probability of state s and $q_{m \cdot s}$ is the probability of receiving message m ("#2 is empty") if state s is true

The solution to the Monty Hall problem is presented in the following table

Note the column 4 (of 5) is the product of columns 2 and 3

Note the column 5 is the ratio of column 4 and $q_m = \sum_{s=1}^{3} j_{s \cdot m}$, which is the number at the bottom of column 4

Computation of posterior probabilities (after message *m*)

State of the world (s)	Prior prob. (π_s)	Likelihood of message *m* ($q_{m \cdot s}$)	Joint prob. (j_{sm})	Posterior prob. ($\pi_{s \cdot m}$)
Prize is in #1	1/3	1/2	1/6	1/3
Prize is in #2	1/3	0	0	0
Prize is in #3	1/3	1	1/3	2/3
			1/2	1.0

Source: Hirshleifer, Jack, and John G. Riley (1992) "The Analytics of Uncertainty and Information," Cambridge (UK): Cambridge University Press, p. 178.

Skewness of returns

The words bullish and bearish make no sense when stock market returns are skewed

If stock market returns are normally distributed—as is typically assumed in asset pricing models—the return distribution is symmetric (around the mean)

We know that, compared to the normal distribution, stock market returns have fat tails, especially on the downside

Sharp stock market declines (crashes) are more common than suggested by the normal distribution while, at the same time, there are no comparable price movements on the upside.

The figure below shows an example of an investor who thinks that the market—most likely—will go up, yet the investor shorts the market, which makes the terms bullish and bearish meaningless

Table 6.2

Event	*Probability*	*Outcome*	*Expectation*
Market goes up	70%	Up 1%	0.7
Market goes down	30%	Down 10%	−3.00
		Total	−2.3

Source: Taleb (2001, p. 88).

Rare events and the peso problem

The probability of rare and extreme events is hard to gauge, simply because they have no precedent

Rare events might be of such extreme consequences that they cast a shadow on today's asset prices and make markets look flawed (i.e., make assets look mispriced) in spite of their low probability of happening

Such an economic situation is called a peso problem, a term that is frequently attributed to Nobel laureate Milton Friedman in comments he made about the Mexican peso market of the early 1970s

In the early 1970s, the exchange rate of the peso was pegged vis-à-vis the U.S. dollar, as it had been since 1954

At the same time, the interest rate on Mexican bank deposits exceeded the interest rate on comparable U.S. bank deposits

This looked like a flaw, since investors could borrow at the low interest rate in the United States, convert dollars into pesos, deposit the pesos in Mexico and at maturity convert back the investment into U.S. dollars at the same exchange rate

The interest rate differential signaled concerns of investors about a devaluation of the peso, which—while appearing unlikely—would be extreme

The skeptical investors were vindicated when the peso lost 46 percent of its value vis-à-vis the U.S. dollar after it was allowed to float.

Regime switching

Relations that seem stable during “normal times” might change in times of crisis

This is particularly important for risk management strategies, such as VAR (value-at-risk)

Such risk management strategies are developed during normal times for times of crisis

In times of crisis, the relations these risk management strategies rely on might not exist, or might look very different

> “During the last five or so years, value-at-risk (VAR) has become an accepted standard in the financial industry. It forms the basis for determining a bank’s regulatory capital for market risk … This approach … assumes that the future movements in risk factors are similar to past movements … Correlation patterns and variances, however, are not stationary, especially when market prices move dramatically. Factors that might exhibit low levels of correlation or association most of the time appear to be highly correlated in volatile times. When the value[s] of nearly all asset classes are moving in lockstep, diversification is not helpful in reducing risk. The actual realized correlation patterns appear to be close to 1. In these times, the volatility of profit and losses will be far greater than VAR would predict.”
>
> Myron R. Scholes (2000, “Crisis and Risk Management,” *American Economic Review, Papers and Proceedings*, Vol. 90, p. 17-21).

The Lucas critique

Nobel laureate Robert Lucas put forward an argument in the 1970s that Frank Knight (1921, III.VII.8) described as follows

> "We must infer what the future situation would have been without our interference, and what change will be wrought in it by our action."

Julian E. Barnes (2001, "A Bicycling Mystery: Head Injuries Piling Up," *The New York Times*, July 29, 2001, http://www.nytimes.com) reports a change in the data-generating process after bicyclists were encouraged to wear helmets

If the risk attitude of bicyclists had not changed, the introduction of helmets would have reduced the number of head injuries

Rather, the absolute and the per-mile numbers of head injuries <u>increased</u>, possibly because helmet-wearing bikers started riding in a more risky manner due to a false perception of safety

In fact, the data-generating process changed <u>because</u> we started making it part of (health) risk management.

[Blank page]

Introduction

2. Expected Utility (Normative Analysis of Decision-Making)

References:

Neumann, John von, and Oskar Morgenstern (1944) *Theory of Games and Economic Behavior*, (2nd ed. 1947), Princeton (NJ): Princeton University Press.

Milgrom, Paul, and John Roberts (1992) *Economics, Organization and Management*, Englewood Cliffs (NJ): Prentice-Hall.

Varian, Hal R. (1992) *Microeconomic Analysis*, 3rd ed. New York: Norton.

The classic way of modeling decisions under uncertainty is the expected utility approach

The axiomatic foundations of expected utility were developed by John von Neumann (1903-1957) and Oskar Morgenstern (1902-1976)

Expected utility is a normative approach in that it predicts human behavior under the presumption of rationality

Actual human behavior might differ from the behavior predicted by expected utility

Note that behavior different than predicted by expected utility need not be irrational

In fact, the notion of irrational behavior is nihilistic because it does not offer an explanation for why the world is the way it is

A constructive approach to explaining seemingly irrational behavior is the concept of bounded rationality, devised by the psychologist Herbert A. Simon (1916-2001), who was honored with the 1978 Nobel Prize in economics.

Expected utility property

Assume that an individual faces a lottery with two possible outcomes:

Wealth x with probability p $(0 < p < 1)$ and wealth y with probability $1-p$

The expected utility property reads:

$$u(\,p \circ x \oplus (1-p) \circ y\,) = p \cdot u(x) + (1-p) \cdot u(y)$$

where $p \circ x \oplus (1-p) \circ y$ is the lottery.

<u>The utility an individual draws from the lottery equals the expected value of the utilities of the outcomes</u>.

Example

Imagine an individual is offered the following deal:

> \$50 of wealth for certain (the “safe project”), or a lottery with equally likely outcomes of \$25 and \$75 of wealth (the “risky project”)

Note that both projects have the <u>same expected value</u>

According to the expected utility property, the individual draws the following utility from the lottery:

$$u\left(\frac{1}{2} \circ \$25 \oplus \frac{1}{2} \circ \$75 \right) = \frac{1}{2} \cdot u(\$25) + \frac{1}{2} \cdot u(\$75)$$

We know that the individual draws the utility $u(\$50)$ from having a wealth of \$50 for certain

If the individual is indifferent about the certain payment and the lottery, she is <u>risk-neutral</u>:

$$u(\$50) = \frac{1}{2} \cdot u(\$25) + \frac{1}{2} \cdot u(\$75)$$

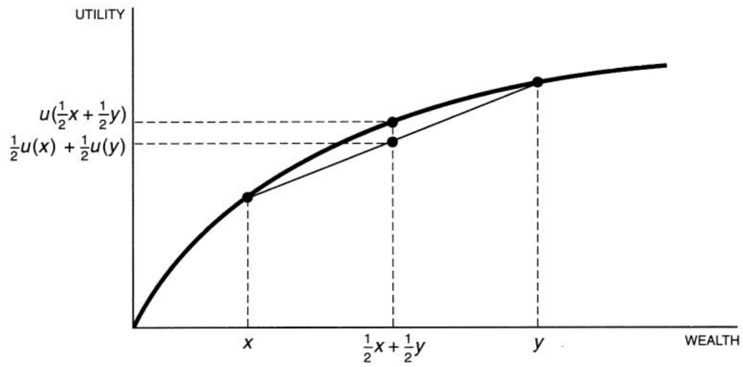
If the individual prefers the certain payment (the “safe project”), she is <u>risk-averse</u>:

$$u(\$50) > \frac{1}{2} \cdot u(\$25) + \frac{1}{2} \cdot u(\$75)$$

If the individual prefers the lottery, this person is <u>risk-seeking</u>:

$$u(\$50) < \frac{1}{2} \cdot u(\$25) + \frac{1}{2} \cdot u(\$75).$$

Graphical illustration of risk aversion



Source: Varian (1992, p. 178).

The vertical difference between the utility of the safe outcome, $u(\frac{1}{2}x + \frac{1}{2}y)$, and the utility of the lottery, $\frac{1}{2}u(x) + \frac{1}{2}u(y)$, is due to a risk premium

The risk premium equals the horizontal distance between the straight and the concave lines at $\frac{1}{2}u(x) + \frac{1}{2}u(y)$

Degrees of risk aversion

As the above figure shows, risk aversion results from the concavity of the expected utility function, $\frac{\partial^2 u(w)}{\partial w^2} := u''(w)$

Based on this concavity, two measures of degrees of risk aversion have been defined by Arrow (1970) and Pratt (1964):

Arrow-Pratt measure of absolute risk aversion:

$$r(w) = \frac{-u''(w)}{u'(w)}$$

Arrow-Pratt measure of relative risk aversion:

$$\tilde{r}(w) = \frac{-w \cdot u''(w)}{u'(w)}$$

Constant relative risk aversion (which implies declining absolute risk aversion) is a reasonable assumption in many applications

Frequently, the assumption of constant absolute risk aversion is made to keep the math tractable

- For marginal changes in wealth, this assumption may be sufficiently accurate.

Certainty Equivalent

The certainty equivalent is the amount of wealth generated by the safe outcome that the individual views as equally valuable as the lottery

For instance, in the above example the individual might have the following preferences:

$$u(\$30) = \frac{1}{2} \cdot u(\$25) + \frac{1}{2} \cdot u(\$75)$$

The certainty equivalent is \$30, because this is the amount of wealth of the safe outcome that gives the individual the same utility as the lottery.

Calculus offers a straightforward approximation to the certainty equivalent, which is helpful in many applications

Let $u(x)$ be an individual's expected utility function; define $\hat{x}$ as the wealth x for which the individual's utility equals the utility of the lottery:

$$u(\hat{x}) = \mathrm{E}[u(x)]$$

The wealth $\hat{x}$ is the certain value (e.g., \$30) that the individual regards as equivalent to the lottery, which generates an uncertain (stochastic) wealth x (e.g., \$25 or \$75, each with probability 1/2).

According to Taylor's theorem, the value of the expected utility function at any point z, $\mathrm{E}[z] = \bar{x}$, equals:

$$u(z) = u(\bar{x}) + (z - \bar{x})\, u'(\bar{x}) + \frac{1}{2}(z - \bar{x})^2\, u''(\bar{x}) + R(z)$$

Setting $R(z)$ equal to zero leads to a second-order Taylor-series approximation:

$$u(z) \approx u(\bar{x}) + (z - \bar{x})\, u'(\bar{x}) + \frac{1}{2}(z - \bar{x})^2\, u''(\bar{x})$$

Applying the expected value operator yields:

$$\mathrm{E}[u(z)] \approx u(\bar{x}) + \mathrm{E}[(z-\bar{x})]\, u'(\bar{x}) + \frac{1}{2}\mathrm{E}[(z-\bar{x})^2]\, u''(\bar{x})$$

$$\Leftrightarrow \ \mathrm{E}[u(z)] \approx u(\bar{x}) + \frac{1}{2}\mathrm{E}[(z-\bar{x})^2]\, u''(\bar{x}) \qquad (1)$$

because of $\mathrm{E}[(z-\bar{x})] = 0$, by definition.

Similarly, we can employ Taylor's theorem at the point $z = \hat{x}$ (at the certain wealth):

$$u(\hat{x}) = u(\bar{x}) + (\hat{x}-\bar{x})\, u'(\bar{x}) + \frac{1}{2}(\hat{x}-\bar{x})^2\, u''(\bar{x}) + R(\hat{x})$$

Applying a <u>first-order</u> Taylor-series approximation yields

$$u(\hat{x}) \approx u(\bar{x}) + (\hat{x}-\bar{x})\, u'(\bar{x}) \quad (2)$$

By definition, the following holds for the certain wealth, $\hat{x}$:

$$\mathrm{E}[u(x)] = u(\hat{x})$$

Inserting Eq. (1) into the right-hand side, and Eq. (2) into the left-hand side of the above equality, yields:

$$u(\bar{x}) + \frac{1}{2}\mathrm{E}[(x-\bar{x})^2]\, u''(\bar{x}) \approx u(\bar{x}) + (\hat{x}-\bar{x})\, u'(\bar{x})$$

$$\Leftrightarrow \; \hat{x}-\bar{x} \approx \frac{1}{2}\frac{u''(\bar{x})}{u'(\bar{x})}\; \mathrm{E}[(x-\bar{x})^2]$$

$$\Leftrightarrow \; \bar{x}-\hat{x} \approx \frac{1}{2} r(\bar{\mathrm{x}})\, \mathrm{Var}[x]$$

The term $r(\bar{x}) = -\dfrac{u''(\bar{x})}{u'(\bar{x})}$ is the Arrow-Pratt measure of (absolute) risk aversion at the expected value of wealth, $\bar{x}$

The term $\mathrm{E}[(x-\bar{x})^2] = \mathrm{Var}[x]$ is the variance of wealth, x

The term $\bar{x}-\hat{x}$ is the risk premium

The risk premium is the <u>difference between the expected value of the lottery and the certainty equivalent</u>.

Inserting the numbers of the aforementioned numerical example yields:

$$\bar{x} = \frac{1}{2} \cdot \$25 + \frac{1}{2} \cdot \$75 = \$50$$

$$\hat{x} = \$30 \text{ (original assumption)}$$

$$\bar{x} - \hat{x} = 20 \text{ (risk premium)}$$

$$\text{Var}[x] \equiv \text{E}[(x - \bar{x})^2] = \frac{1}{2} \cdot (\$25 - \$50)^2 + \frac{1}{2} \cdot (\$75 - \$50)^2 = \$^2 \ 625$$

Degree of (absolute) Arrow-Pratt risk aversion:

$$r(x) = \frac{2(\bar{x} - \hat{x})}{\text{Var}[x]} = \frac{2 \cdot \$20}{\$^2 \ 625} = 0.0624/\$$$

Degree of relative Arrow-Pratt risk aversion:

$$\tilde{r}(x) = \hat{x} \cdot r(x) = 30 \cdot 0.0624 = 1.92$$

Note that we used $\hat{x}$ as a reference point.

In summary, we derived a price for risk (the risk premium) by determining the compensation the risk-averse individual demands for bearing risk:

$$\bar{x} - \hat{x} \,\hat{\approx}\, u(\text{E}[x]) - \text{E}[u(x)] \equiv u(\text{expected value for certain}) - u(\text{lottery})$$

[Blank page]

Introduction

3. Heuristics and Biases (Descriptive Analysis of Decision-Making I)

Reference:

Tversky, Amos, and Daniel Kahneman (1974) "Judgement under Uncertainty: Heuristics and Biases," *Science* 185, 1124-31, reprinted in: Daniel Kahneman, Paul Slovic, and Amos Tversky (1982), eds., *Judgement under Uncertainty: Heuristics and Biases*. Cambridge (UK): Cambridge University Press, 3-20.

The study of decision-making address both normative and descriptive questions

The normative analysis of decision-making is concerned with the nature of rationality and the logic of decision-making

Rational decision-making (under uncertainty) is modeled in the expected utility approach developed by von Neumann and Morgenstern; see the chapter "Expected Utility (Normative Analysis of Decision-Making)."

The descriptive analysis is concerned with people's preferences and beliefs as they are, not as they should be under the presumption of consistent (i.e., rational) behavior

This chapter and the chapter "Choice, Values and Frames (Descriptive Analysis of Decision-Making II)" deal with the descriptive analysis of decision-making.

Quantitative finance theory—outside the sphere of Behavioral Finance—rests on the expected utility approach.

Heuristics

People rely on a limited number of heuristic principles that reduce the complex task of assessing probabilities and predicting values to simpler judgmental operations

In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors (biases)

There are three heuristics people use to assess probabilities and to predict values

Representativeness

Availability

Anchoring.

1. Representativeness

When answering questions of the following types, people typically rely on the representativeness heuristic:

What is the probability that object A belongs to class B?

What is the probability that event A originates from process B?

What is the probability that process B will generate event A?

In answering these questions, people evaluate probabilities by the degree to which A is representative of B or, in other words, by the degree to which A resembles B

For instance, when A is highly representative of B, the probability that A originates from B is judged to be high

On the other hand, if A is not similar to B, the probability that A originates from B is judged to be low.

For an illustration of judgement by representativeness, consider an individual who is described as follows:

> "Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail."

How do people assess the probability that Steve is engaged in a particular occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)?

In the representativeness heuristic, the probability that Steve is a librarian, for example, is assessed by the degree to which he is representative of, or similar to, the stereotype of a librarian, a finding that has been confirmed in experiments.

The representative heuristic leads to serious errors that fall into six categories

i. Insensitivity to prior probability (Bayesian prior) of outcomes

One of the factors that has no effect on representativeness but should have a major effect on probability is the prior probability, or base-rate frequency, of the outcomes

In the case of Steve, for example, the fact that there are many more farmers than librarians in the population should enter into the estimate of the probability that Steve is a librarian rather than a farmer

Consideration of the base-rate frequency, however, does not affect the similarity of Steve to the stereotype of librarians and farmers

Thus, if people evaluate probability by representativeness, prior probabilities will be neglected.

For instance, in an experiment, subjects were told that a person randomly drawn from a population of 70 (30) engineers and 30 (70) lawyers has the following personality description:

> "The person is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues."

Note that his description is pure noise for it reveals no information relevant to question of whether the person is an engineer or a lawyer

The subjects in the experiment judged the probability of the person being an engineer to be 0.5, regardless of whether the stated proportions of engineers in the group was 0.7 or 0.3

On the other hand, when no personality description was offered, the probabilities estimated corresponded to the base-rate frequencies.

In conclusion, when no specific evidence is given, prior probabilities are utilized properly; when worthless evidence (noise) is given, prior probabilities are ignored.

ii. Insensitivity to sample size

To evaluate the probability of obtaining a particular result in a sample drawn from a specified population, people typically apply the representativeness heuristic

Consequently, the judged probability of a sample statistic is essentially independent of sample size.

Consider the following example

"A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower.

For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?

The larger hospital (21)

The smaller hospital (21)

About the same (that is, within five percent of each other) (53)"

- The values in parentheses are the number of graduate students that chose each answer
  - Thus, most subjects judged the probability of obtaining more than 60 percent boys to be the same in the small and in the large hospital, presumably because these events are equally representative of the general population
    - In contrast, sampling theory entails that the expected number of days on which more than 60 percent of the babies are boys is much greater in the small hospital than in the large one, because a large sample is less likely to stray from 50 percent.

iii. Misconceptions of chance

People expect that a sequence of events generated by a random process represents the essential characteristics of that process even when the sequence is short

For instance, in considering tosses of a coin for heads (H) and tails (T), people erroneously regard the sequence

H-T-H-T-H-T

to be more likely than the sequence

H-H-H-T-T-T

which does not appear random, and also more likely than the sequence

H-H-H-H-T-H

which does not represent the fairness of the coin.

Another consequence of the belief in local representativeness is the gambler's fallacy

After observing a long run of red on the roulette wheel, for example, some people erroneously believe that black is now due, presumably because the occurrence of black will result in a more representative sequence than the occurrence of an additional red

Chance is commonly viewed as a self-correcting (mean-reverting) process in which a deviation in one direction induces a deviation in the opposite direction to restore the equilibrium

In fact, a random process—as it unfolds—does not "correct" deviations, it merely dilutes them.

iv. Insensitivity to predictability

Predictions, for instance on the corporate earnings, are often made by representativeness

For example, suppose one is given a description of a company and is asked to predict its future profit

If the description of the company is very favorable, a very high profit will appear most representative of that description

If the description is mediocre, a mediocre performance will appear most representative.

Typically, the degree to which the description is favorable is unaffected by the reliability of that description or by the degree to which it permits accurate prediction

Hence, if people predict solely in terms of the favorableness of the description, their predictions will be insensitive to the reliability of the evidence and to the expected accuracy of the prediction.

This mode of judgement violates the normative statistical theory in which the extremeness and the range of predictions are controlled by considerations of predictability

<u>When predictability is nil, the same prediction should be made in all cases</u>

For example, if the descriptions of companies provide no information relevant to earnings, then the same value (e.g., for earnings per share) should be predicted for all companies

On the other hand, the higher the predictability, the wider the range of predicted earnings should be in cross section.

In one of the studies on predictions by representativeness, subjects were given descriptions of student teachers during a practice lessen

- One group of subjects was asked to <u>evaluate</u> the quality of the described practice lesson, the other group was asked to <u>predict</u>—using the same scoring mechanism—the standing of the student teacher after 5 years into the job
  - The judgements made under the two conditions were identical!
    - That is, the prediction of a remote criterion (success of a teacher after 5 years) was identical to (and thus as extreme as) the evaluation of the information on which the prediction was based (the quality of the practice lesson).

v. The illusion of validity

As we have seen, people often predict by selecting the outcome (for example, an occupation) that is most representative of the input (for example, the description of a person)

The confidence people have in their prediction depends primarily on the degree of representativeness (that is, on the quality of the match between the selected outcome and the input) with little or no regard for the factors that limit predictive accuracy (illusion of validity)

Thus, people express great confidence in the prediction that a person is a librarian when given a description of his personality that matches the stereotype of librarians, even if the description is scanty, unreliable, or outdated.

The internal consistency of a pattern of inputs is a major determinant of one's confidence in predictions based on these inputs

For instance, people express more confidence in predicting the final grade-point average (GPA) of a student whose first-year record consists entirely of B's than in predicting the GPA of a student whose first-year record includes many A's and C's

Highly consistent input patterns are frequently observed when the input variables are highly redundant or correlated

Hence, people have great confidence in predictions based on redundant (correlated) input variables (i.e., in input variables that add no information yet add to the internal consistency of the input pattern) that do not improve the accuracy of the prediction.

vi. Misconception of regression to the mean (mean reversion)

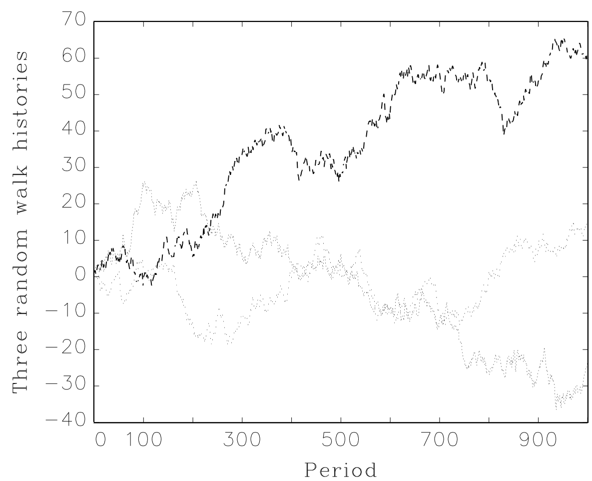
Variables may follow a random walk (with or without drift) or—alternatively—revert (regress) to the mean (which might be stationary or a trend)

A simple example of a random walk (with drift $\mu = 0$ and standard deviation σ) reads:

$$x_{t+1} = x_t + \varepsilon \ , \ \varepsilon \sim N(0;\sigma)$$

The best prediction of next period's value of the variable is the value of the current period

A variable that follows a random walk does not possess a finite variance; it can wander off without bounds (see figure below)

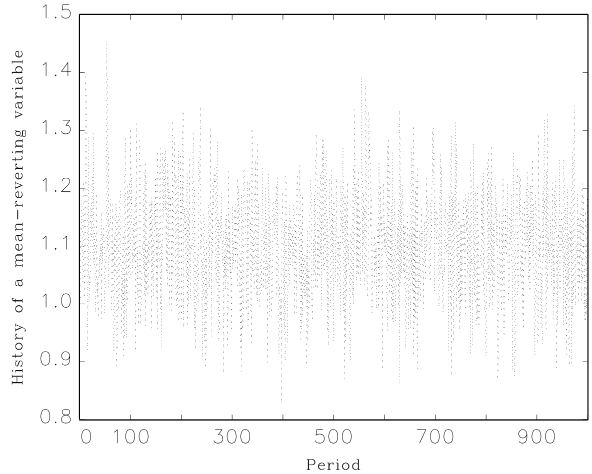


A simple example of a mean-reverting process reads:

$$x_{t+1} = \bar{x} + \rho \cdot (\bar{x} - x_t) + \varepsilon \ , \ 0<\rho<1 \ , \ \varepsilon \sim N(\mu;\sigma^2)$$

A mean-reverting variable is called stationary (or trend-stationary, if it reverts to a trend growth rate)

The variable follows a self-correcting process where the degree of correction is determined by the parameter ρ



For some variables (e.g., the rate of inflation), it is difficult to judge whether they are stationary, while for other variables (e.g., on-the-job performance) it is obvious

On-the-job performance has no potential of being unbounded.

Often times, people are not aware of mean reversion as the following example illustrates

In a discussion of flight training, experienced instructors noted that praise for an exceptionally smooth landing is typically followed by a poorer landing on the next try, while harsh criticism after a rough landing is usually followed by an improvement on the next try

The instructors erroneously concluded that verbal rewards are detrimental to learning, while verbal punishments are beneficial

The conclusion is unwarranted because performance is a mean-reverting variable.

2. Availability

There are situations in which people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind

For instance, one may evaluate the probability that a given business venture will fail by imagining various difficulties it could encounter.

Availability is a useful clue for assessing frequency or probability, because instances of large classes (events more available) are usually reached better and faster than instances of less frequent classes (events less available)

However, availability is affected by factors other than frequency and probability, which might introduce biases in predictions that rely on availability

i. Biases due to the retrievability of instances

When the size of a class is judged by the availability of its instances, a class whose instances are easily retrieved appear more numerous than a class of equal frequency whose instances are less retrievable

In an experiment, subjects heard a list of well-known personalities of both sexes and were subsequently asked to judge whether the list contained more names of men than of women; different lists were presented to different groups of subjects

In some of the lists the men were relatively more famous than the women, and in others the women were more famous than the men

In each of the lists, the subjects erroneously judged that the class (sex) that had the more famous personalities was the more numerous.

In addition to familiarity, there are other factors, such as salience, that affect the retrievability of instances

> For example, the impact of seeing a house burning on the subjective probability of such accidents is probably greater than the impact of reading about a fire in the local paper.

Furthermore, recent occurrences are likely to be relatively more available than earlier occurrences

> It is a common experience that the subjective probability of traffic accidents rises temporarily when one sees a car overturned by the side of the road.

ii. Biases due to the effectiveness of a search set

Suppose one samples a word (of three letters or more) at random from an English text. Is it more likely that the word starts with *r* or that *r* is the third letter?

People approach this problem by recalling words that begin with *r* (road) and words that have *r* in the third position (car) and assess the relative frequency by the ease with which words of the two types come to mind.

Because it is much easier to search for words by their first letter than by their third letter, most people judge words that begin with a given consonant to be more numerous than words in which the same consonant appears in the third position

People do this even for consonants, such as *r*, that are more frequent in third position than in the first.

iii. Biases of imaginability

Imaginability plays a role in the evaluation of probabilities in real-life situations

The risk involved in an adventurous expedition, for instance, is evaluated by imagining contingencies with which the expedition is not equipped to cope

If many such difficulties are vividly portrayed, the expedition can be made to appear exceedingly dangerous, although the ease with which disasters are imagined need not reflect their actual likelihood

Conversely, the risk involved in an undertaking may be grossly underestimated if some possible dangers are either difficult to conceive of, or simply do not come to mind.

iv. Illusory correlation

In an experiment, naïve judges were confronted with information concerning several hypothetical mental patients

The data for each patient consisted of a clinical diagnosis and a drawing of a person that this patient had made.

Later the judges estimated the frequency with which each diagnosis (such as paranoia or suspiciousness) had been accompanied by various features of the drawing (such as peculiar eyes)

The subjects markedly overestimated the frequency of co-occurrence of natural associates, such as suspiciousness and peculiar eyes

This illusory correlation effect was extremely resistant to contradictory data and even persisted when the correlation between symptom and diagnosis was negative.

Illusory correlation is due to availability

The judgement of how frequently two events co-occur may be based on the strength of the associative bond between them

For instance, the illusory correlation between suspiciousness and peculiar drawing of the eyes might be due to the fact that suspiciousness is more readily associated with the eyes than with any other part of the body.

3. Anchoring

In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer

The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation

In either case, adjustments are typically insufficient or, in other words, different starting points yield different estimates, which are biased toward the initial values (anchoring).

Anchoring leads to biases that fall into three categories

i. Insufficient adjustment

In an experiment, subjects were asked to estimate various quantities, for instance, the percentage of African countries in the United Nations

A wheel of fortune was spun in front of the subject; the wheel generated a number that appeared random to the subject but was actually preset

Different groups of subjects were given different numbers.

Then the subject was asked to indicate whether the number on the wheel is lower or higher than the quantity (e.g., the percentage of African countries in the UN) they were asked to estimate

Next, the subject was asked to estimate the quantity he was asked for by moving upward or downward from the number given by the wheel

It turned out that the median estimates of the percentage of African nations in the UN were 25 and 45 for groups that received 10 and 65, respectively, as starting points from the wheel.

Anchoring also occurs when subjects base their estimates on incomplete computations

In an experiment, high school students were asked to estimate a product of numbers within five seconds

Students in one group were given the product

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

while students in another group were given the product

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

To answer such question quickly, people tend to perform a few steps of computation and estimate the product by extrapolation or adjustment

Because adjustments are typically insufficient, this procedure should lead to underestimation

Furthermore, because the result of the first few steps of multiplication (performed from the left to the right) is higher in the descending sequence than in the ascending sequence, the former expression should be judged larger than the latter.

The experiment confirmed both predictions: The median estimate for the ascending order was 512, while the median estimate for the descending sequence was 2,250; the correct answer is 40,320.

ii. Biases in evaluation of conjunctive and disjunctive events

In an experiment, subjects were given the opportunity to bet on one of two events

Three types of events were used

Simple events

Drawing a red marble from an urn that contains 50 percent red and 50 percent black marbles ($p = 0.50$).

Conjunctive events

Drawing seven red marbles in a row (with replacement) from an urn that contains 90 percent red marbles and 10 percent black marbles ($p = 0.48$).

Disjunctive events

Drawing a red marble at least once in seven successive tries (with replacement) from an urn that contains 90 percent black marbles and 10 percent red marbles ($p = 0.52$).

A significant majority of subjects preferred to bet on the conjunctive event ($p = 0.48$) rather than the simple event ($p = 0.50$)

Subjects also preferred to bet on the simple event ($p = 0.50$) rather than on the disjunctive event ($p = 0.52$)

This pattern of choice illustrates the general finding that people tend to overestimate the probability of conjunctive events and underestimate the probability of disjunctive events.

Bias in the evaluation of compound (i.e., conjunctive) events are particularly prevalent in the context of planning

The successful completion of an undertaking, such as the development of a new product, typically has a conjunctive character in that for the undertaking to succeed, each of a series of events must occur

Even when each of these events is very likely, the overall probability of success can be quite low if the number of events is large

The general tendency to overestimate the probability of conjunctive events leads to unwarranted optimism in the evaluation of the likelihood that a plan will succeed or that a project will be completed on time.

Conversely, disjunctive structures are typically encountered in the evaluation of the likelihood of rare (catastrophic) events

A complex system, such as a nuclear reactor or a human body, will malfunction if any of its essential components fails

Even when the likelihood of failure in each component is slender, the probability of an overall failure can be high if many components are involved

Because of anchoring, people tend to underestimate the probabilities of failure in complex systems.

In summary, the direction of the anchoring bias can sometimes be inferred from the structure of the event

The chain-like structure of conjunctions leads to overestimation, while the funnel-like structure of disjunctions leads to underestimation.

iii. Anchoring in the assessment of subjective probability distributions

In decision analysis, experts are often required to express their beliefs about a quantity, such as the value of the Dow Jones Industrial Average (DJIA) on a particular day, in the form of a probability distribution

Such a distribution is usually constructed by asking the person to select values of the quantity that correspond to specified percentiles of his subjective probability distribution

For instance, the expert may be asked to select a number, $X_{0.95}$, such that his subjective probability that the DJIA will be lower than this number is 95 percent

Along with a value $X_{0.05}$ which the expert expects the DJIA to exceed with a probability of 95 percent, we obtain a 90 percent confidence interval for the DJIA.

Experimental studies have shown that about 30 percent of the actual realizations of a variable fall outside a person's subjective 98 percent confidence interval of this variable, which means that people's subjective probability distributions are not well calibrated

That is, the subjects state confidence intervals that are overly narrow or, in other words, reflect more certainty than is justified by the subjects' knowledge about the assessed quantities

The bias is common to naïve and to sophisticated subjects, and is not eliminated by introducing scoring rules that provide incentives for external calibration.

The bias of delivering overly narrow confidence intervals is—at least partly—due to anchoring

- When individuals are asked for a confidence interval, e.g., $\{X_{0.05};X_{0.95}\}$ for the DJIA, individuals tend to start out with a point estimate (e.g., X_{50}) and then adjust this value upward (for $X_{0.95}$) and downward (for $X_{0.05}$)
  - Such adjustments—as discussed above (wheel of fortune experiment)—tend to be insufficient, resulting in an overly narrow confidence interval.

In support of the anchoring hypothesis, it can be shown that subjective probabilities are systematically altered by a procedure in which one's own point estimate does not serve as an anchor

- In an experiment, a group of subjects were either asked for their X_{10} or X_{90} assessment for 24 different quantities (such as the air distance from New Delhi to Beijing)
- Another group of subjects received the median X_{10} (or X_{90}) judgement of the first group and was asked to assess the odds that this value exceeds (or—in the case of X_{90}—falls short of) the true value of the quantity (e.g., air distance from New Delhi to Beijing)
  - In the absence of any bias, the second group should retrieve the odds specified by the first group, i.e., 9:1
  - However, if even odds (i.e., fifty-fifty) or the stated value serve as an anchor for the second group, the odds of the second group should be less extreme (i.e., closer to 1:1)
  - Indeed, the median odds, stated by the second group across all 24 problems, were 3:1.

When the judgement of the two groups were tested for external calibration, it was found that subjects of the first group were too extreme, in accord with earlier studies

- The events that the first group defined as having a probability of 10 percent actually obtained in 24 percent of the cases
- In contrast, subjects in the *second group* were too conservative
  - Events to which they assigned an average probability of 34 percent were obtained in 26 percent of the cases.

Summary

We discussed three heuristics that people employ in making judgements under uncertainty

Representativeness, which is usually employed when people are asked to judge the probability that an object or event A belongs to class or process B

Availability of instances or scenarios, which is often employed when people are asked to assess the frequency of a class or the plausibility of a particular development

Anchoring, which usually applies in numerical prediction when a relevant value is available.

The three heuristics are highly economical and usually effective, but they lead to systematic and predictable errors

The biases are not restricted to laymen; experienced researcher are also prone to these errors when they think intuitively.

[Blank page]

Introduction

4. Choice, Values, and Frames (Descriptive Analysis of Decision-Making II)

Reference:

Kahneman, Daniel, and Amos Tversky (1984) "Choices, Values, and Frames," *American Psychologist* 39, 341-50, reprinted in: Daniel Kahneman and Amos Tversky (2000), eds., *Choices, Values, and Frames*. Cambridge (UK): Cambridge University Press, 1-16.

Abstract of Kahneman and Tversky (1984/2000)

> "We discuss the cognitive and the psychological determinants of choice in risky and riskless contexts. The psychophysics of value induce risk aversion in the domain of gains and risk seeking in the domain of losses. The psychophysics of chance induce overweighting of sure things and of improbable events, relative to events of moderate probability. Decision problems can be described or framed in multiple ways that give rise to different preferences, contrary to the invariance criterion of rational choice. The process of mental accounting, in which people organize the outcomes of transactions, explains some anomalies of consumer behavior. In particular, the acceptability of an option can depend on whether a negative outcome is evaluated as a cost or as an uncompensated loss. The relation between decision values and experience is discussed."

Risky choice

The analysis of risky choice deals with the cognitive and psychophysical factors that determine the value of risky prospects

(Psychophysics is concerned with the relationship between physical stimulation and sensation, for instance, between the physical intensity of a stimulus and the psychological intensity of the sensation it causes.)

The paradigmatic example of decision under risk is the acceptability of a gamble (lottery) that yields monetary outcomes with specified probabilities.

The psychological approach to decision-making has its roots in an essay by Daniel Bernoulli (1700-1782), published in 1738

Bernoulli attempted to explain why people are generally averse to risk and why risk aversion decreases with increasing wealth

To illustrate risk aversion, compare …

… a gamble that offers an 85 percent chance of winning $1,000 and a 15 percent chance of winning nothing (statistical expected value: $0.85 \times \$1000 + 0.15 \times \$0 = \$850$)

… with a sure profit of $800

Most people (in most situations) prefer the sure outcome to the "positive" gamble, although the gamble offers the higher (statistical) expected payoff

In general, preferring a sure outcome to a gamble with higher or equal expected value is a manifestation of risk aversion

Similarly, a rejection of the sure outcome in favor of a gamble with equal or lower expected value (but a possible outcome that exceeds the sure payoff) is a manifestation of risk-seeking.

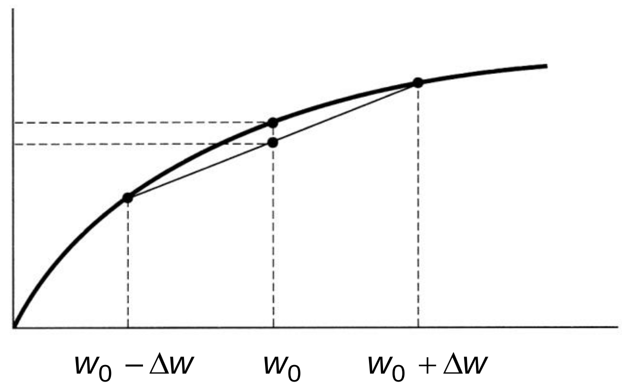
It is customary in decision analysis to describe the outcomes (of gambles) in terms of total wealth, w

A risk-averse individual prefers current wealth to a fifty-fifty chance of moving up or down $\Delta w = 20$ from her initial wealth w_0

Technically, this means that the value a risk-averse individual attaches to a lottery is a <u>concave function</u> of the wealth outcomes

$$w_0 \succ 0.5 \times (w_0 - \Delta w) + 0.5 \times (w_0 + \Delta w)$$

("$\succ$" means "preferred to")



Psychological evidence suggests that people do not normally think in terms of total wealth but in terms of gains and losses

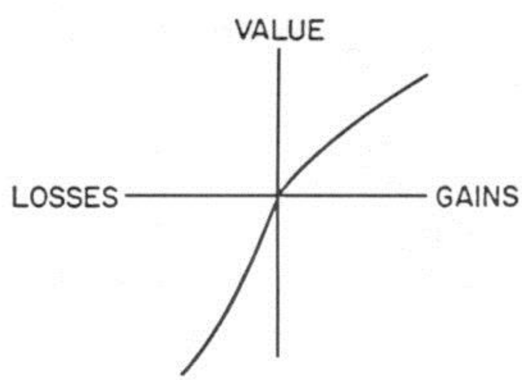
The assumption that the effective carriers of subjective value are changes of wealth rather than ultimate states of wealth is central to prospect theory, a concept developed by Kahneman and Tversky (1979, "An Analysis of Decision under Risk," *Econometrica* 47, 263-91)

According to prospect theory, subjective value is a concave function of the size of a gain

The same generalization applies to losses as well

The difference in subjective value between a loss of $200 and a loss of $100 appears greater than the difference in subjective value between a loss of $1,200 and a loss of $1,100

When the value functions for gains and losses are pieced together, we obtain an S-shaped function of the typed displayed in the following figure



Source: Kahneman and Tversky (2000, p. 3)

The function displayed in the above figure…

- …evaluates gains and losses rather than total wealth
- …is concave in the domain of gains and convex in the domain of losses
- …is considerably steeper for losses than for gains—a manifestation of loss aversion.

Loss aversion implies that a loss of $x is more repellent than a gain of the same amount is attractive

- Loss aversion explains people's reluctance to bet on a fair coin for equal stakes
  - The attractiveness of a possible gain is not nearly sufficient to compensate for the aversion to the possible loss
    - Kahneman and Tversky (1984/2000) report that most respondents in a sample of undergraduates refused to stake $10 on the toss of a coin if they stood to win less than $30

While risk aversion has played a central role in modeling decision-making since Bernoulli, risk-seeking—as implied by the convexity of the value of losses—is an innovation of prospect theory

- Indeed, there is empirical evidence that risk-seeking is a robust effect, in particular when the probabilities of loss are substantial
  - Consider, for instance, a situation in which an individual is forced to choose between an 85 percent chance of losing $1,000 (with a 15 percent chance of losing nothing) and a sure loss of $800
    - A large majority of people prefers the "negative" gamble to the sure loss
      - Preferring the gamble to the sure loss is risk-seeking behavior, because the expected value of the gamble (-$850) is lower than the sure loss.

The S-shaped value function, which implies risk-seeking in losses and risk aversion in gains, violates at least two principles that are generally accepted as being necessary for rational (i.e., consistent) human behavior

- Dominance
  - Dominance demands that if prospect A is at least as good as prospect B in at least one respect, then A should be preferred to B.
- Invariance
  - Invariance requires that the preference order between prospects should not depend on the manner in which they are described.

Framing of outcomes

Risky prospects can be described as gains and losses relative to the status quo or asset positions that incorporate initial wealth

Invariance requires that changing the description of outcomes (framing) does not alter the preference order

The following pair of problems illustrates a violation of this requirement

The total number of respondents in each problem is denoted by *N*, and the percentages of subjects that chose the available options are indicated in parentheses

Problem 1 ($N = 152$): Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If program A is adopted, 200 people will be saved. (72 percent)
If program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved. (28 percent)

Which program would you favor?

The formulation of problem 1 adopts as a reference point a state of affairs in which the disease is allowed to take 600 lives

The outcomes of the programs include the reference state and two possible gains measured by the number of lives saved

A large majority of respondents shows risk-averse preferences in that it prefers saving 200 lives for sure to a gamble that offers a one-third chance of saving 600 lives.

Now consider another problem in which the same story is followed by a different description of the prospects associated with the two programs:

Problem 2 ($N = 155$): Imagine that the U.S. …

If program C is adopted, 400 people will die. (22 percent)
If program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die. (78 percent)

Which program would you favor?

It is easy to verify that the set of options C and D in problem 2 is identical to the set of options A and C in problem 1

Unlike problem 1, which is phrased in terms of <u>gains</u>, problem 2 is phrased in terms of <u>losses</u>

When options are phrased in terms of losses, people tend to exhibit risk-seeking preferences; indeed, the majority chooses the gamble in problem 2.

Respondents confronted with their conflicting answers are typically puzzled

Even as they are rereading the problems, the respondents still wish to be risk-averse in the "lives saved" version and risk-seeking in the "lives lost" version

At the same time, the respondents state that they wish to obey invariance and give consistent answers.

The following pair of problems elicits preferences that violate the dominance principle of rational choice

Problem 3 ($N = 86$): Choose between:

E. 25 percent chance of winning $240 and a 75 percent chance of losing $760 (0 percent)

F. 25 percent chance of winning $250 and a 75 percent chance of losing $750 (100 percent)

It is easy to see that in problem 3, F dominates E, and all respondents choose accordingly.

Now consider the following problem:

Problem 4 ($N = 150$): Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate which options you prefer.

Decision (i) Choose between:

A. A sure gain of $240 (84 percent)

B. A 25 percent chance of a $1,000 gain and a 75 percent chance of a gaining nothing (16 percent)

Decision (ii) Chose between:

C. A sure loss of $750 (13 percent)

D. A 75 percent chance of losing $1,000 and a 25 percent chance of losing nothing (87 percent)

As expected from the previous analysis, a large majority of the subjects made a risk-averse choice in preferring the sure gain to the gamble in the first decision, and a large majority of subjects made a risk-seeking choice in preferring the gamble to the sure loss in the second decision.

In fact, 73 percent of the subjects chose “A and D,” and only 3 percent chose “B and C”—a finding not revealed by the numbers given above

- Preferring “A and D” to “B and C”—in a situation of simultaneous choice ("First examine both decisions…")—means <u>rejecting a dominated conjunction</u>
  - “A and D” yields a 25 percent chance of winning $240 and a 75 percent chance of losing $760
    - This corresponds to option E in problem 3
  - “B and C,” on the other hand, yields a 25 percent chance of winning $250 and a 75 percent chance of losing $750
    - This corresponds to option F in problem 3
- Taken together, the susceptibility of choice to framing, along with the S-shaped value function leads to <u>violation of dominance</u> in a set of concurrent decisions.

The psychophysics of chances

Expected utility—devised by Daniel Bernoulli—assumes that the value (utility) of an uncertain prospect (lottery) is obtained by adding the utility levels of the possible outcomes, each weighted by its probability

$$u(\, p \circ x \oplus (1-p) \circ y \,) \;=\; p \cdot u(x) + (1-p) \cdot u(y)$$

where $p \circ x \oplus (1-p) \circ y$ is the lottery

Imagine that you are given a ticket to a lottery that has a single prize of $300

How does the value of the ticket vary as a function of the probability of winning the prize?

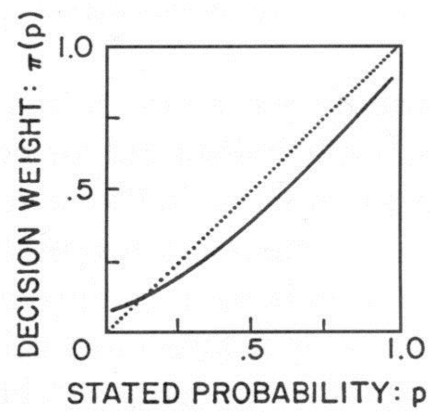
According to expected utility, the value of the ticket is a linear function of the probability of winning the prize, p:

$$u(\, p \circ \$300 \oplus (1-p) \circ \$0 \,) \;=\; p \cdot u(\$300) + (1-p) \cdot u(0)$$

Psychological evidence, on the other hand, suggests that the value of a lottery <u>is not</u> a linear function of the probability of winning, p

An increase from 0 percent to 5 percent or from 95 percent to 100 percent seems to have a larger effect than an increase from 30 percent to 35 percent

In other words, a change from impossibility to possibility or from possibility to certainty has a bigger impact than a comparable change in probability in the middle of the scale—a <u>category boundary effect</u> in decision weights (see figure below)



Source: Kahneman and Tversky (2000, p. 7)

The weighting function displayed above reflects the following empirical findings about decision weights

- The convexity of the weighting function implies that decision weights are <u>regressive</u> with respect to stated probabilities
  - Save for the neighborhoods of the endpoints, an increase of 5 percentage points in the probability of winning increases the value of the prospect of winning by less than 5 percentage points.
- Decision weights are lower than the corresponding probabilities over most of the range
  - Underweighting of moderate and high probabilities relative to sure outcomes contributes to risk aversion in gains by reducing the attractiveness of "positive" gambles; underweighting also contributes to risk-seeking in losses by attenuating the aversion of "negative" gambles

Low probabilities, however, are overweighted quite grossly or neglected altogether, making the decision weights highly unstable in that region

The overvaluation of low probabilities enhances the value of long shots and amplifies the aversion to a small chance of a severe loss

Consequently, people are often risk-seeking in dealing with improbable gains and risk-averse in dealing with unlikely losses

This explains the attractiveness of lottery tickets (unlikely gains) and insurance policies (unlikely losses).

The nonlinearity (convexity) of decision weights inevitably leads to violation of invariance, as illustrated in the following pair of problems

Problem 5 ($N = 85$): Consider the following two-stage game. In the first stage, there is a 75 percent chance of dropping out of the game without winning anything and a 25 percent chance of moving on to the second stage. If you reach the second stage, you have a choice between:

- **A.** A sure gain of $30 (74 percent)
- **B.** An 80 percent chance of winning $45 (26 percent)

Your choice must be made before the game starts, i.e., before the outcome of the first stage is known. Please indicate which option you prefer.

Problem 6 ($N = 81$): Which of the following options do you prefer?

- **C.** A 25 percent chance of winning $30 (42 percent)
- **D.** A 20 percent chance of winning $45 (58 percent)

Because there is a 25 percent chance of moving on to the second stage in problem 5, option A is equivalent to option C in terms of <u>probabilities and outcomes</u>; the same holds true for options B and D

However, the <u>preferences</u> are not the same in the two versions (which were played with real money)

In problem 5, the majority prefers option A, while in problem 6 the majority prefers option D—a striking violation of invariance

Violation of invariance has been confirmed in many settings, among them one with human lives as outcomes.

Kahneman and Tversky (1984/2000) attribute violation of invariance to the interaction of two factors: nonlinearity of the weighting function and framing

- More specifically, Kahneman and Tversky propose that in problem 5, people ignore the first stage of the game—a violation of expected utility—and focus on the second stage (problem of framing)
  - It has been reported that the choices observed in problem 5 resemble the choices made when the first stage of the game is not mentioned
    - In other words, the $30 outcome in the two-stage game of problem 5 is treated as if it were certain—a phenomenon called <u>pseudo-certainty effect</u>
    - Because a sure outcome is overweighted in comparison with outcomes of moderate or high probability, option A in problem 5 is comparatively more attractive vis-à-vis option B than is option C in problem 6 vis-à-vis option D (problem of nonlinearity).

The nonlinearity (convexity) of decision weights makes probabilistic insurance appear unattractive

Suppose you are indifferent to buying earthquake insurance at the going rate or, in other words, the going rate is "just acceptable"

As you hesitate to sign the policy, your insurance agent suggests the following deal

"For half the rate, I offer you a policy that insures against earthquakes that happen on odd days, which means that you are covered for more than half the days of the year."

Most people find such probabilistic insurance unattractive

As the weighting function displayed in the above figure suggests—starting anywhere in the low probability region—reducing the probability from p to $p/2$ is worth less than reducing the probability from $p/2$ to zero

Consequently, for bearing half the risk that you are willing to eliminate at the margin, your agent has to offer you a discount of more than 50 percent.

The aversion to probabilistic insurance is significant for three reasons

- According to expected utility, probabilistic insurance should be preferred to normal insurance when the latter is just acceptable
  - This is because the marginal utility from eliminating risk is decreasing.
- Second, probabilistic insurance represents many forms of protective action, such as medical checkups, high-quality tires—goods and services that typically reduce the probability of a hazard without eliminating it
  - The aversion to probabilistic insurance affects the prices of these goods and services in the marketplace.
- Third, the acceptability of insurance can be manipulated by the framing of the contingencies
  - An insurance policy that covers fire but not flood, for instance, could be evaluated either as <u>elimination</u> of risk of property loss from fire or as <u>reduction</u> of risk of property loss from natural disasters overall
    - Imagine a vaccine that reduces the probability of contracting a disease from 20 percent to 10 percent
      - There is experimental evidence that this vaccine is valued as less attractive when it is described as effective in half of the cases than when it is described as fully effective against one of two exclusive and equally dangerous virus strains.

Transactions and trades

In situations of choice, there are three categories of mental accounting that people might employ to evaluate the gains and losses that are associated with the options they face

Consider the following situation of choice

Problem 7: Imagine that you are about to purchase a jacket for $125 and a calculator for $15. The salesman informs you that the calculator you wish to buy is on sale for $10 at the other branch of the store, located 20 minutes away. Would you make the trip?

[68 percent of $N = 88$ respondents were willing to drive to the other branch to save $5 on the $15 calculator.]

A minimal account regards only the difference among possible choices and disregards the features they share

Driving to the other store is framed as a gain of $5

A topical account relates the consequences of possible choices to a reference level that is determined by the context within which the decision arises

The relevant topic is the purchase of the calculator, and the benefit of the trip is thus framed as a reduction of the price from $15 to $10

Because the potential saving is associated only with the calculator, the price of the jacket is not included.

A comprehensive account evaluates the gains and losses of possible choices in relation to broader categories—for instance, monthly expense.

The finding that 68 percent of $N = 88$ respondents were willing to drive to the other branch to save $5 on the calculator does not tell about the category of mental accounting that people tend to apply

Kahneman and Tversky (1984/2000) suggest that in the described situation of choice people spontaneously frame decisions in terms of topical accounts

To test this hypothesis, Kahneman and Tversky constructed a version of the test in which they interchanged the prices of the two goods

Problem 8: Imagine that you are about to purchase a jacket for $15 and a calculator for $125. The salesman informs you that the calculator you wish to buy is on sale for $120 at the other branch of the store, located 20 minutes away. Would you make the trip?

[29 percent of $N = 93$ respondents were willing to drive to the other branch to save $5 on the $125 calculator.]

The sharp decline in the fraction of respondents that are willing to make the trip—to 29 percent from the previous 68 percent— supports the notion of topical organization of accounts, for problems 7 and 8 are identical in terms of a minimal and a comprehensive account

The topical organization of mental accounts leads people to evaluate gains and losses in relative rather than absolute terms, resulting in large variations in the rate at which money is exchanged for things.

Now consider an example where the posting of a cost to an account is controlled by topical organization

Problem 9: ($N = 200$) Imagine that having decided to see a play, you paid $10 for a ticket. As you arrive at the theater, you discover that you have lost the ticket. The seat is not numbered, and the ticket cannot be recovered.

Would you pay $10 for another ticket?

Yes (46 percent) No (54 percent)

Problem 10: ($N = 183$) Imagine that having decided to see a play, you are willing to pay $10 for a ticket. As you arrive at the theater, you discover that you have lost a $10 bill.

Would you still pay $10 for a ticket for the play?

Yes (88 percent) No (12 percent)

Why are so many people unwilling to spend another $10 after having lost a ticket, if they would readily spend that sum after losing an equivalent amount of cash?

Kahneman and Tversky (1984/2000) attribute the difference to topical organization of mental accounts

Going to the theater is viewed as a transaction in which the cost of the ticket is exchanged for the experience of seeing a play

Buying a second ticket increases the cost of seeing the play, possibly to a level that exceeds their willingness-to-pay

In contrast, the loss of the cash is not posted to the account of the play, and it affects the purchase of a ticket only by making the individual slightly less wealthy.

Losses and payments

Many decisions take the form of a choice between retaining the status quo and accepting an alternative that is advantageous in some respects and disadvantageous in others

The status quo might serve as a reference level for all attributes of the alternatives, which are evaluated in terms of gains and losses relative to the status quo

Because losses weigh more heavily than gains, the decision-maker is biased in favor of retaining the status quo

The reluctance of people to give up assets that come with the status quo is called endowment effect

Note that when it is more painful to give up an asset than it is pleasurable to obtain it, the individual's willingness-to-pay (WTP) for the asset is lower than the willingness-to-accept (WTA) giving up the asset

The endowment leads to discrepancies between buying and selling prices, and to reluctance to trade

For instance, in an experiment subjects were assigned hypothetical jobs that differed in salary (S) and in temperature (T) at the workplace

The respondents were asked to imagine that they held a particular position (S_1,T_1), and then were offered the option of moving to an alternative position (S_2,T_2), which was better in one respect and worse in another

It was found that most subjects that were assigned to the position (S_1,T_1) did not wish to move to (S_2,T_2), and most subjects that were assigned to the position (S_2,T_2) did not want to switch to (S_1,T_1).

The endowment effect and the previously discussed loss aversion are related phenomena

Loss aversion was discussed above in the context of risky prospects (decisions under uncertainty), whereas the endowment was found in situations of trade where no risk was involved

Because losses loom larger than gains, framing becomes a powerful tool in manipulating people's purchasing decisions

For instance, an insurance premium can be framed as a sure loss (which appeals to WTA) or a price for avoiding the possibility of a larger loss (which appeals to WTP)

Consider the choice between a sure loss of $50 and a 25 percent chance of losing $200—a risky prospect of identical expected value

It has been reported that 80 percent of the subjects confronted with this situation of choice preferred the gamble to the sure loss—exhibiting risk-seeking behavior

On the other hand, only 35 percent of the subjects refused to pay $50 to insure themselves against a 25 percent chance of losing $200

Thus, the modal preferences reversed when the same amount of money ($50) was not framed as an uncompensated loss but as a price for avoiding the risk of an even greater loss.

The following two problems address a similar case in the domain of gains (that is, risky prospects with positive, rather than negative expected value)

> **Problem 11**: Would you accept a gamble that offers a 10 percent chance of winning $95 and a 90 percent chance of losing $5?
>
> **Problem 12**: Would you pay $5 to participate in a lottery that offers a 10 percent chance of winning $100 and a 90 percent chance of winning nothing?

- A total of 132 subjects answered the two questions, which were separated by a short filler; half the subjects received the questions in the reverse order
- Although the two problems offer identical options, 55 of the respondents expressed different preferences in the two versions
  - Among these 55 subjects, 42 rejected the gamble in problem 11 but accepted the equivalent lottery in problem 12
    - Thinking of $5 as a payment makes the venture more acceptable than thinking of the same amount as a loss.

If negative outcomes are more acceptable when framed as payments than when framed as losses, people might fall prey to a <u>dead-loss effect</u>

- In literature, a case has been discussed where a man who develops tennis elbow shortly after paying the membership fee in a tennis club keeps playing in agony to avoid wasting his investment
  - Continuing to play maintains the evaluation of the membership fee as a price paid rather than a dead loss.

Introduction

5. The Sharpe-Lintner CAPM

References:

Brealey, Richard A., and Stewart C. Myers (2000) *Principles of Corporate Finance*, 6th ed. New York: McGraw-Hill, Ch. 7, 8, 9.

The Sharpe-Lintner Capital Asset Pricing Model (CAPM) is one of the cornerstones of asset pricing and capital budgeting

We start out by studying the simple mean-variance model of portfolio choice with one safe and two risky financial assets

Note that if the return on an asset is normally distributed, the mean and the variance of the return is a sufficient statistic

Also, remember that we showed in the foregoing chapter that the mean-variance approach may be viewed as a Taylor-series approximation to the expected utility function.

In a first step toward the mean-variance model of portfolio choice we bundle the two risky assets in a portfolio, which we call asset 1

The return on a portfolio, p, that consists of the safe asset (asset 0) and the risky asset (asset 1) equals:

$$r_p = x_0 \cdot r_0 + (1 - x_0) \cdot r_1$$

where x_0 is the fraction of (initial) wealth invested in the risk-free asset

The expected (value of the) return on portfolio p reads:

$$E[r_p] = x_0 \cdot r_0 + (1 - x_0) \cdot E[r_1]$$

The expected excess return on the portfolio (the return in excess of the return on the risk-free asset) thus equals:

$$E[r_p] - r_0 = (x_0 - 1) \cdot r_0 + (1 - x_0) \cdot E[r_1]$$

$$\Leftrightarrow E[r_p] - r_0 = (1 - x_0) \cdot (E[r_1] - r_0)$$

The variance of the return on portfolio p reads:

$$\text{Var}[r_p] = (1 - x_0)^2 \cdot \text{Var}[r_1]$$

$$\Leftarrow \text{Std}[r_p] = (1 - x_0) \cdot \text{Std}[r_1]$$

where $\text{Std}[x] \equiv \sigma_x \equiv \sqrt{\text{Var}[x]}$ for any random variable x

Consequently, we obtain:

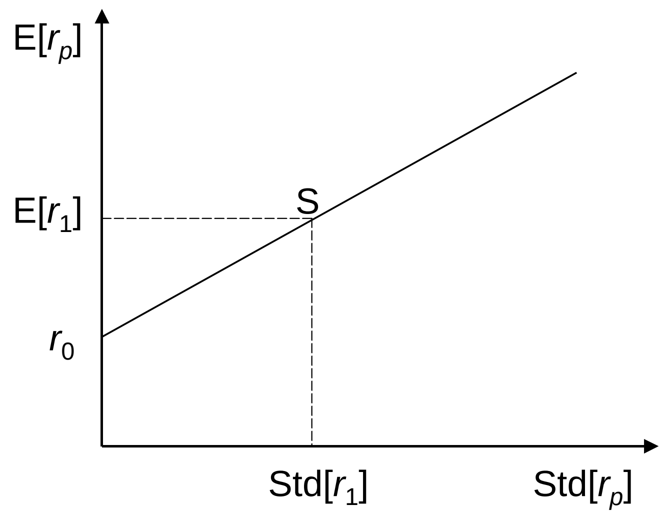
$$1 - x_0 = \frac{\text{Std}[r_p]}{\text{Std}[r_1]}$$

The fraction of wealth invested in the risky asset equals the standard deviation of the portfolio return normalized by the standard deviation of the return on the risky asset

Inserting this ratio into the above equation for the portfolio’s expected excess return yields:

$$E[r_p] = r_0 + \frac{E[r_1] - r_0}{\text{Std}[r_1]} \cdot \text{Std}[r_p]$$

There is a linear relationship between the portfolio's expected return, E $[r_p]$, and the portfolio's risk as measured by the standard deviation of returns, Std $[r_p]$



The straight line in the figure represents all combinations of risk and expected return that are available to the investor

If the investor keeps all her wealth in the risk-free asset, her expected (and actual) return equals r_0, and the standard deviation of returns, Std $[r_p]$, as well as the expected excess return, E $[r_1] - r_0$, are equal to zero

If the investor holds the risky asset only, her expected return equals E $[r_1]$, and the standard deviation of returns amounts to Std $[r_1]$

All combinations of risk and return to the southwest of point S represent portfolios with positive amounts invested in the risk-free asset.

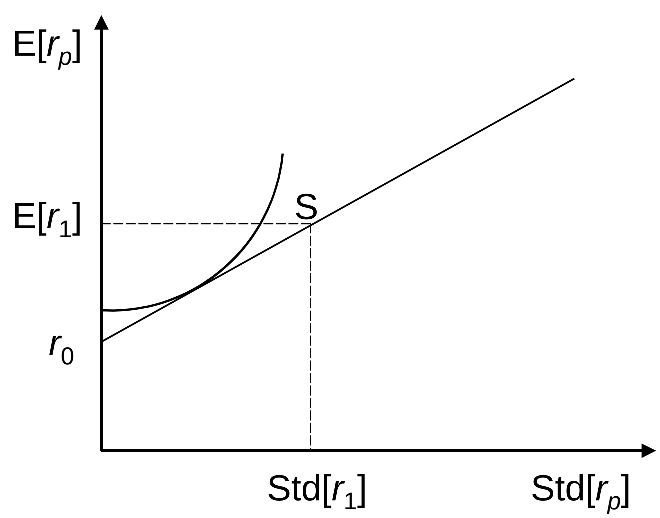
Leveraged portfolios are portfolios located to the northeast of point S

The investor leverages up the portfolio by borrowing (at the risk-free rate) and investing in the risky asset

For instance, if the investor borrows an amount twice her initial wealth she props up her expected excess return to 300% of E $[r_1]$, but also faces 300% of the risk level of Std $[r_1]$.

The portfolio choice depends on the investor's preferences for risk and return

Every investor holds a portfolio that consists of the risky assets and (negative, positive or zero amounts of) the risk-free asset



Note that higher indifference lines represent higher levels of utility.

In a <u>second step</u>, we break up the risky asset into two risky assets, which allows for <u>diversification</u>

The return on the portfolio now reads:

$$r_p = x_0 \cdot r_0 + x_1 \cdot r_1 + x_2 \cdot r_2 \ , \quad \sum_{i=0}^{2} x_i = 1$$

$$\Leftrightarrow \ r_p = (1 - x_1 - x_2) \cdot r_0 + x_1 \cdot r_1 + x_2 \cdot r_2$$

The expected return equals:

$$\mathrm{E}[r_p] = (1 - x_1 - x_2) \cdot r_0 + x_1 \cdot \mathrm{E}[r_1] + x_2 \cdot \mathrm{E}[r_2]$$

Thus we obtain for the expected excess return:

$$\mathrm{E}[r_p] - r_0 = x_1 \cdot \mathrm{E}[r_1 - r_0] + x_2 \cdot \mathrm{E}[r_2 - r_0]$$

The variance of the portfolio returns reads:

$$\begin{aligned}
\mathrm{Var}[r_p] &= \mathrm{Cov}[r_p, r_p] \\
&= \mathrm{Cov}[(1 - x_1 - x_2) \cdot r_0 + x_1 \cdot r_1 + x_2 \cdot r_2 \ , \ r_p] \\
&= (1 - x_1 - x_2) \cdot \mathrm{Cov}[r_0, r_p] + x_1 \cdot \mathrm{Cov}[r_1, r_p] + x_2 \cdot \mathrm{Cov}[r_2, r_p] \\
&= \quad x_1 \cdot \mathrm{Cov}[r_1 \ , \ (1 - x_1 - x_2) \cdot r_0 + x_1 \cdot r_1 + x_2 \cdot r_2] \\
&\quad + x_2 \cdot \mathrm{Cov}[r_2 \ , \ (1 - x_1 - x_2) \cdot r_0 + x_1 \cdot r_1 + x_2 \cdot r_2] \\
&= \quad (x_1)^2 \cdot \mathrm{Var}[r_1] + x_1 \cdot x_2 \cdot \mathrm{Cov}[r_1, r_2] \\
&\quad + (x_2)^2 \cdot \mathrm{Var}[r_2] + x_1 \cdot x_2 \cdot \mathrm{Cov}[r_1, r_2] \\
&= \quad (x_1)^2 \cdot \mathrm{Var}[r_1] + (x_2)^2 \cdot \mathrm{Var}[r_2] + 2 \cdot x_1 \cdot x_2 \cdot \mathrm{Cov}[r_1, r_2]
\end{aligned}$$

It can be shown that if (and only if) the returns on the two risky assets are not perfectly correlated with each other, i.e., if the correlation coefficient between the returns, $-1 \leq \frac{\text{Cov}[r_1, r_2]}{\text{Std}[r_1] \cdot \text{Std}[r_2]} \leq 1$, is smaller than 1, investors gain from diversification

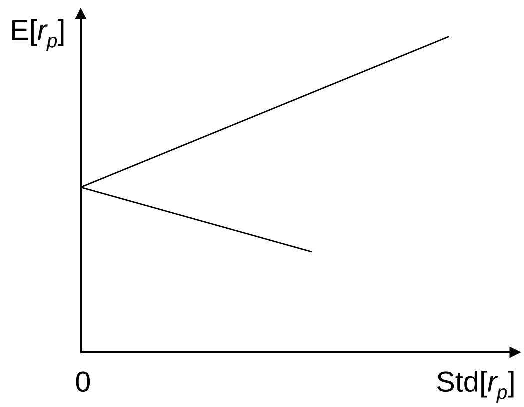
Diversification means to keep portfolio risk low by spreading the wealth over a multitude of risky assets whose returns are not perfectly correlated

As an example, imagine you own an orchard, and apples and grapes are the only fruits you grow

The following table shows the payoff structure of both fruits for sunny and rainy years, the only two possible meteorological outcomes

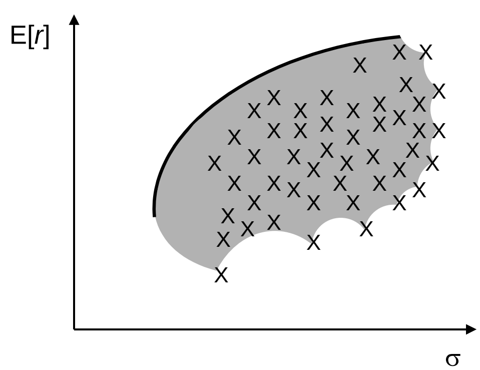
Weather / Payoff	Apples	Grapes
Sunny	80	120
Rainy	100	80

The following figure presents all combinations between risk and expected return for alternative allocations of the two risky assets for a given amount of wealth



The southeastern endpoint is the "all apples" portfolio, while the northeastern endpoint is the "all grapes" portfolio.

For a <u>multitude of risky assets</u>, this figure looks as follows



Assets (and portfolios, i.e., complex assets) that lie on the thick line are <u>mean-variance efficient</u>

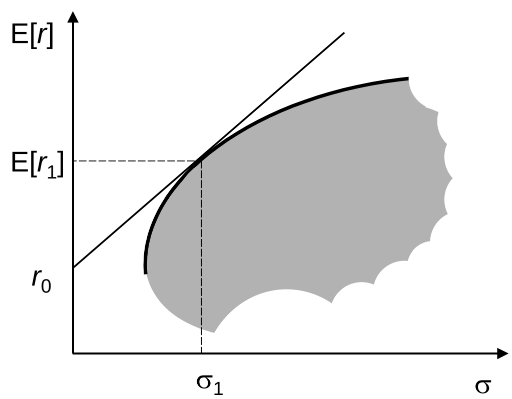
An asset (a portfolio) is mean-variance efficient if there is no other asset (portfolio) of equal expected return [equal risk] that has lower risk [higher expected return].

Above, we established the following relationship for the expected return on a portfolio that consists of the risk-free asset (asset 0) and the risky asset (asset 1):

$$E[r_p] = r_0 + \frac{E[r_1] - r_0}{\mathrm{Std}[r_1]} \cdot \mathrm{Std}[r_p]$$

Asset 1 comprises the domain of risky assets, the market portfolio or, simply, the market

We know (from the figure below) that all combinations of the risk-free asset and the market are efficient, i.e., there are no portfolios of the same expected return (same risk) that offer lower risk (higher return)



How do we price an asset in the market portfolio (the gray area), given that we can hold any combination between the risky asset the market portfolio (that is, any risk-return combination located on the straight line)?

Given that we can hold any combination of the risk-free asset and the market portfolio, the risk of an (arbitrary) asset j in the market portfolio is priced only to the extent that it resembles such a combination

Which combination of the risk-free asset and the market portfolio replicates (an arbitrary) risky asset j most closely?

In finding this replicating portfolio, we look for the portfolio on the straight line that has the smallest sum of squared deviations in returns relative to the risky asset in question

The optimization problem for determining the replicating portfolio reads

$$\underset{\tilde{x}_0}{\text{Min}}\ \text{Var}[r_j - \tilde{x}_0 \cdot r_0 - (1 - \tilde{x}_0) \cdot r_1]$$

The first order-condition is given by

$$\frac{\partial\, \text{Var}[r_j - \tilde{x}_0 \cdot r_0 - (1 - \tilde{x}_0) \cdot r_1]}{\partial\, \tilde{x}_0} = 0$$

with

$$\mathrm{Var}[r_j - \tilde{x}_0 \cdot r_0 - (1-\tilde{x}_0)\cdot r_1]$$

$$= \mathrm{Var}[r_j] - 2\cdot \mathrm{Cov}[r_j\ ,\ \tilde{x}_0 \cdot r_0 + (1-\tilde{x}_0)\cdot r_1]$$

$$+\ \mathrm{Var}[\tilde{x}_0 \cdot r_0 + (1-\tilde{x}_0)\cdot r_1]$$

$$= \mathrm{Var}[r_j] - 2\cdot \mathrm{Cov}[r_j\ ,\ (1-\tilde{x}_0)\cdot r_1] + \mathrm{Var}[(1-\tilde{x}_0)\cdot r_1]$$

$$= \mathrm{Var}[r_j] - 2\cdot (1-\tilde{x}_0)\cdot \mathrm{Cov}[r_j, r_1] + (1-\tilde{x}_0)^2 \cdot \mathrm{Var}[r_1]$$

we obtain for the first-order condition:

$$\frac{\partial\, \mathrm{Var}[\bullet]}{\partial \tilde{x}_0} = 2\cdot \mathrm{Cov}[r_j, r_1] - 2(1-\tilde{x}_0)\cdot \mathrm{Var}[r_1] = 0$$

$$\Leftrightarrow\ (1-\tilde{x}_0) = \frac{\mathrm{Cov}[r_j, r_1]}{\mathrm{Var}[r_1]} \equiv \beta_j$$

The left-hand side of the equation is the share of the market (asset 1) in the replicating portfolio

The right-hand side is the degree of co-movement of the returns of asset j with the market returns, called the CAPM beta, β_j

> The CAPM beta, as a measure of co-movement with the market portfolio, gauges the systematic risk ("market risk") of asset j

Systematic risk, which is captured by the asset's CAPM beta, is part of the asset's total risk

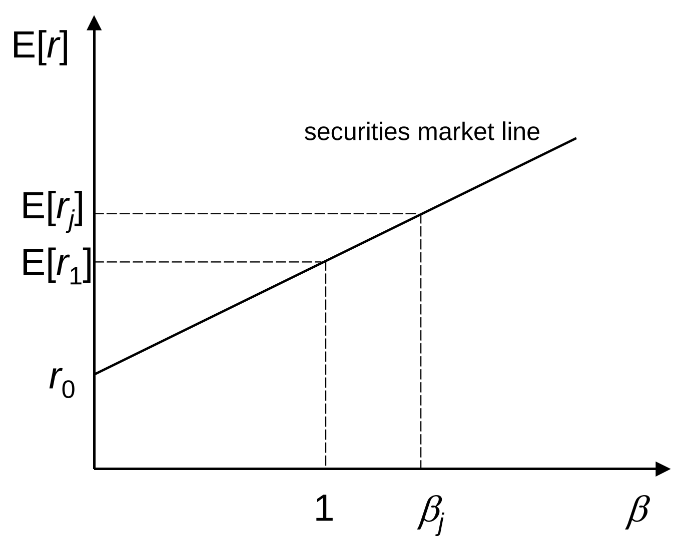
The other component of the asset's total risk is idiosyncratic risk, which is risk that is specific to the asset (i.e., not common to any other asset)

CAPM betas of corporate stocks can be estimated econometrically

Equity betas are published on finance Web sites, such as <http://www.bloomberg.com>

On the Bloomberg Web site, key in the stock's ticker symbol (e.g. "KO" for Coca-Cola Company), and the equity beta is displayed in the last row of "Fundamentals."

The figure below shows the market securities line—the line on which all assets lie in equilibrium



[Blank page]

Introduction

6. Efficient Markets

Reference:

Fama, Eugene F. (1970) "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25, 383-417.

Fama, Eugene F. (1991) "Efficient Capital Markets," *Journal of Finance* 46, 1575-617.

Fama, Eugene F. (1998) "Market Efficiency, Long-term Returns, and Behavioral Finance," *Journal of Financial Economics* 49, 283-306.

Grossman, Sanford J., and Joseph E. Stiglitz (1980) "On the Impossibility of Informationally Efficient Markets," *American Economic Review* 70, 393-408.

Jensen, Michael Grossman, Sanford J., and Joseph E. Stiglitz (1980) "On the Impossibility of Informationally Efficient Markets," *American Economic Review* 70, 393-408.

In 1970, Fama defined three categories of market efficiency

The <u>weak form</u> of market efficiency

Past returns do not predict future returns.

The <u>semi-strong form</u> of market efficiency

All publicly available information is incorporated in asset prices at any time.

The <u>strong form</u> of market efficiency

All private information is incorporated in asset prices at any time.

Insider trading (i.e., trading on information available to firm insiders only) is a violation of the strong form of market efficiency.

Market efficiency *per se* is not testable

Market efficiency must be tested jointly with some model of equilibrium, an asset-pricing model (joint-hypothesis problem)

One can only test whether information is properly reflected in prices in the context of a pricing model that defines the meaning of “properly”

As a result, when one finds anomalous evidence on the behavior of returns, the way the anomaly should be attributed to market inefficiency or a bad model of equilibrium is ambiguous.

The impossibility of efficient markets

The concept of market efficiency (which in its most general form states that security prices reflect all available information) is a concept of <u>static equilibrium</u>

Static equilibrium models, by their very nature, can be solved for states of equilibrium only

Static equilibrium models of asset pricing do not describe the <u>adjustment paths</u> asset prices follow on their way from an original to a new equilibrium.

In a seminal paper, Grossman and Stiglitz (1980) make the point that if asset prices reflect all publicly available information at any time, no investor would have an incentive to collect information and incorporate it into prices

Consequently, in a world with a continuous flow of information, a certain degree of inefficiency must be present at any time to give investors an incentive to collect and filter information and incorporate what is relevant into asset prices.

The critique by Grossman and Stiglitz let to the adoption of a weaker and economically more sensible version of the efficiency hypothesis

Following Jensen (1978), market efficiency implies that security prices reflect information to the point where the marginal benefits of acting on information (the profits to be made) do not exceed the marginal costs.

Behavioral Finance

Behavioral Finance is largely derived from stock market anomalies, i.e., stock market behavior that appears to be inconsistent with rational decision-making and (or) efficient markets

As for as the efficient market aspect goes, the joint-hypothesis problem kicks in every time an anomaly is detected

Also, there is a data-snooping problem

If a significance level of five percent is applied, five false positives will show if we keep searching for anomalies long enough.

Fama (1998) makes a valid point when criticizing research on anomalies (and thus a large branch of Behavioral Finance)

> "Market efficiency survives the challenge from the literature on long-term return anomalies. Consistent with the market efficiency hypothesis that the anomalies are chance results, apparent overreaction to information is about as common as underreaction, and postevent continuation of preevent abnormal returns is about as frequent as postevent reversal. Most important, consistent with the market efficiency prediction that apparent anomalies can be due to methodology, most long-term return anomalies tend to disappear with reasonable changes in technique."

As far as the rational decision-making is concerned, this is where the most significant contribution of Behavioral Finance lies

- Traditional finance does not distinguish between normative and positive statements
  - Starting from a model of rational decision-making (i.e., expected utility) traditional finance derives market outcomes as they would occur if individuals behaved rationally (by expected utility standards)
  - Actual behavior (positive sciences) might differ from prescribed behavior (normative sciences), and Behavioral Finance acknowledges that
  - Differences between actual and prescribed behavior does not imply irrationality
    - Tests of rationality also suffer from the joint-hypothesis problem, as rationality is defined by a benchmark model that might be inappropriate
    - Clearly, humans are subject to bounded rationality as the cognitive and information-processing capacities of the human brain are limited.

Generally, one might want to be conservative in associating stock market behavior that one does not understand with departures from market efficiency or rational decision-making

- Many phenomena man used to view as Acts of God, modern science can explain.

[Blank page]

Introduction

7. Predictability of Stock Returns

Reference:

Hawawini, Gabriel, and Donald B. Keim (1995) “On the Predictability of Common Stock Returns: World-Wide Evidence,” in: R.A. Jarrow, V. Maksimovic and W.T. Ziemba, eds., *Handbooks in Operations Research and Management Science*, Vol. 9: *Finance*, Amsterdam: Elsevier, 497-544.

As mentioned in the chapter “Efficient Markets,” an implication of the strong form of market efficiency is that the (logarithmic) stock price follows a random walk (with upward drift; see below)

If the (log) stock price follows a random walk, the return (i.e., change in the log price) (above and beyond the drift) is purely random and thus unpredictable

In the following we look at empirical evidence of predictability of stock returns

Does evidence for predictability of stock returns imply stock market inefficiency?

The answer is no; for two reasons

First, there is the joint hypothesis problem, mentioned in the chapter “Efficient Markets”

Remember that the null hypothesis in testing stock market efficiency states that “stock markets are efficient and returns behave according to a prescribed equilibrium model.”

Second, the factors that are identified as having predictive power for stock market returns might be risk factors

For instance, the Sharpe-Lintner CAPM allows for drift in the (logarithmic) stock price due to (i) the time value of money (which manifests in the risk-free rate of return) and (i) systematic risk (which manifests in the CAPM beta)

Is the excess return (return above and beyond the risk-free rate) predictable?

It sure is! The higher the CAPM beta, the higher the expected (excess) return!

Consequently, if there are risk factors that are not captured by the CAPM, in an efficient market these risk factors are priced if they cannot be eliminated through diversification

Hence, if there are risk factors beyond what is captured by the applied equilibrium model (e.g., the Sharpe-Lintner CAPM), and if investors demand compensation for bearing these risks, then these risk factors contribute to the expected (excess) returns

When the CAPM is extended to capture these extra risk factors, F_k, the literature speaks about multi-factor models:

$$\mathrm{E}[r_j] = a_0 + a_1 \cdot \beta_j + \sum_{k=1}^{K} a_k \cdot F_k$$

where r_j is the return on the stock of company j and β_j is the company's CAPM beta.

Clearly, there is a circularity in the logic of labeling factors that have predictive power for stock returns "risk factors"

Calling these factors risk factors allows one to reconcile the empirical evidence with the efficient markets paradigm

Yet, the question remains whether the compensation investors receive for the presence of these risk factors is appropriate.

Finance literature has identified <u>four main factors</u> that contribute to the excess return of stocks, i.e., have predictive power for stock returns beyond what is captured by the CAPM beta

- Market capitalization (of the company's stock)
  - All else equal, the stock of smaller companies offers a higher excess return, on average.
- P/E ratio (price-to-earnings ratio)
  - All else equal, the higher the (four-quarter trailing) P/E ratio, the lower the excess return, on average
  - Note that published P/E ratios might be four-quarter forward-looking P/E ratios (rather than four-quarter trailing P/E ratios), or might look back only three quarters and use an estimate for the current quarter
  - Also, there are alternative earnings concepts that are used for calculating P/E ratios
    - Operating earnings (after tax) is net income from on-going operations, which excludes non-recurring items, such as extraordinary items, earnings from discontinued operations, some special items, and gains or losses on sales of assets
    - Ordinary earnings (after tax) is net income from continuing operations calculated using generally accepted accounting principles (GAAP), which includes all current revenue and expenses.
- P/B ratio (price-to-book ratio of the company's stock)
  - All else equal, the lower the price-to-book ratio, the higher the excess return, on average.

Dividend yield (last dividend divided by a quarter of the share price)

All else equal, the higher the dividend yield on the stock over the last four quarters (dividend—if paid—are paid quarterly) of the fiscal year, the higher the excess return, on average.

The table below shows that the P/E ratio, the P/B ratio and the dividend yield exhibit strong pair-wise correlation, given their common factor—the stock price

Table 5

Average rank correlations (*t*-statistics) between several predetermined characteristics for NYSE and AMEX stocks, and also between these characteristics and returns during the following year (1962–1989)

	Market capitalization	Earnings/ price	Price/ book	Dividend yield	Price
Earnings/price	0.05 (1.70)				
Price/book	0.30 (15.09)	−0.29 (−10.99)			
Dividend yield	−0.01 (−0.51)	0.36 (14.50)	−0.47 (−31.03)		
Price per share	0.78 (84.94)	0.11 (4.02)	0.33 (17.40)	−0.13 (−7.12)	
Annual return ($t+1$)	0.03 (0.92)	0.12 (6.09)	−0.07 (−3.45)	0.04 (1.46)	0.04 (1.15)

Correlations are computed annually using ranks of individual stocks. All rankings are conducted at the end of March, using prices at that time and accounting numbers for the previous fiscal year.

Source: Hawawini and Keim (1995, p. 511).

In a seminal paper, Eugene F. Fama and Kenneth R. French (1992, "The cross-section of expected stock returns," *Journal of Finance* 47, 427-66) showed that two factors, size and the price-to-book ratio, are sufficient to characterize cross-sectional differences in expected excess returns.

Most interestingly, the effects of size, P/B ratio and P/E ratio (i) fall almost entirely into the month of January and (ii) covary within the calendar year

The following figure illustrates the covariation of the three effects within the calendar year and their concentration in the month of January

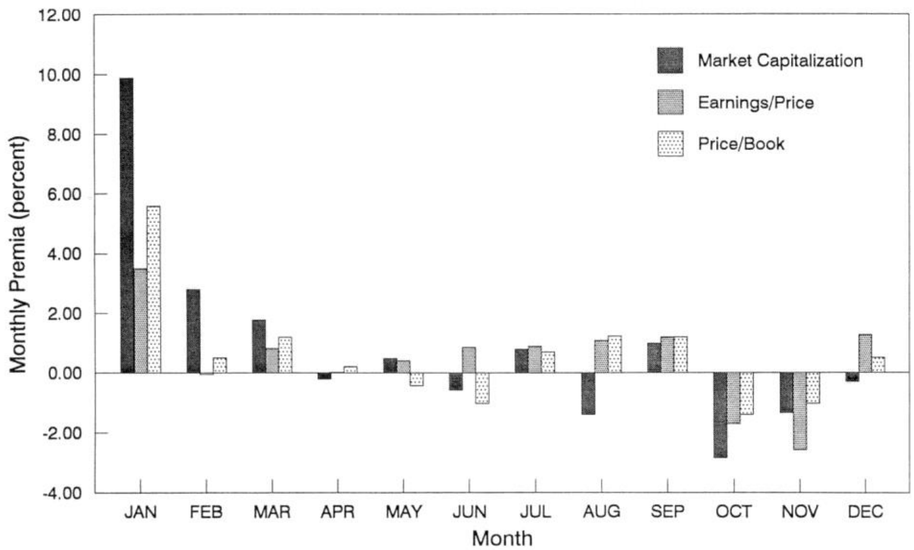


Fig. 1. Monthly difference in returns for extreme deciles (4/62–12/89) (e.g. smallest market cap. – largest market cap.). Based on monthly returns of value-weighed decile portfolios of NYSE and AMEX stocks. Portfolios are (independently) constructed on March 31 of each year using March 31 shares outstanding and prices, and prior-year-end accounting values. Aside from new listings and delistings, which are added to or dropped from the portfolio as they occur during the year, the portfolio composition remains constant over the following twelve months. The portfolios contain only December 31 fiscal closers.

Source: Hawawini and Keim (1995, p. 512).

The January effect is largely due to tax-induced selling

If a stock has suffered a severe loss over the calendar year, personal investors might realize the tax loss late in the year before repurchasing the stock after a period of at least 30 days

Given that smaller companies exhibit less institutional ownership (and a higher fraction of direct ownership by personal investors), the tax-induced selling (late in the calendar year) and buying (in January) is more pronounced with small caps (small capitalization stocks) than large caps.

The following table shows that the P/E and P/B effects are not only strongly correlated within the calendar year (see figure above) but also strongly correlated long-term

The following figure also shows that the size effect correlates little with the other two effects, long-term

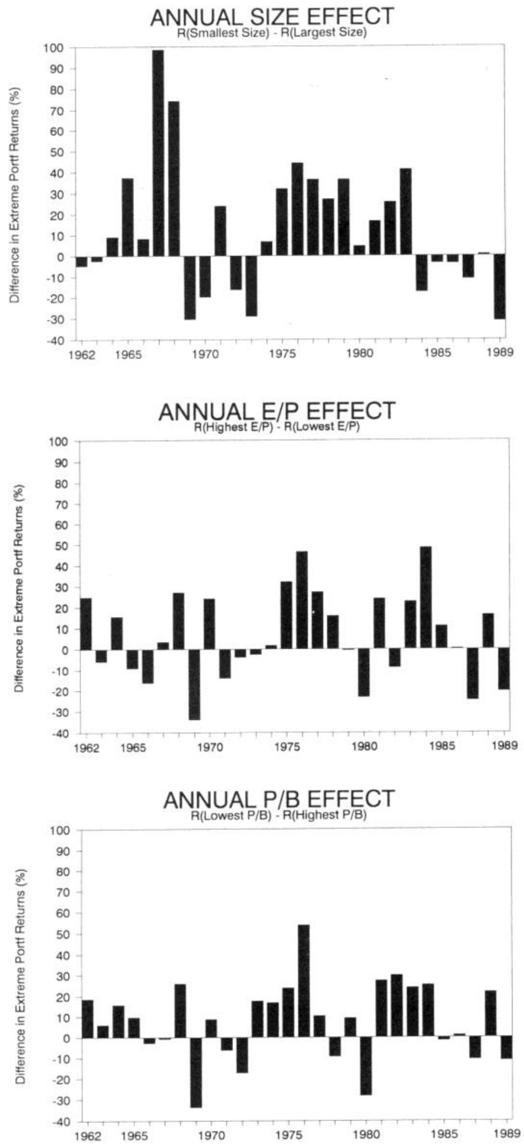


Fig. 2. Annual size, E/P and P/B effects over period 1962–1989. Based on monthly returns of value-weighed decile portfolios of NYSE and AMEX stocks. Portfolios are (independently) constructed on March 31 of each year using March 31 shares outstanding and prices, and prior-year-end accounting values. Aside from new listings and delistings, which are added to or dropped from the portfolio as they occur during the year, the portfolio composition remains constant over the following twelve months. The portfolios contain only December 31 fiscal closers.

Source: Hawawini and Keim (1995, p. 512).

There are seasonal patterns other than the January effect that have been evidenced in empirical studies

Most important is the weekend effect that documents significant positive returns on Friday and significant negative returns on Monday.

Also, there is evidence for serial correlation both in the return of individual stocks and stock market indices

In subsequent chapters on stock market overreaction and stock market underreaction we discuss the causes of serial correlation in the returns of individual stocks.

Conclusion

- It is widely accepted that the Sharpe-Lintner CAPM does not fully capture the risk factors that bear on the expected return of stocks
- Most important are the Fama-French factors (market capitalization and price-to-book ratio)
- Predictability of stock market returns does not imply stock market inefficiency due to the joint-hypothesis problem
- Also, what held true in the past might not hold true in the future
  - If the predictability in stock market returns suffices to earn abnormal returns (i.e., returns in excess of what is warranted by the risk-profile of the portfolio), investors (e.g., hedge funds) will act on it and eliminate the predictability.

[Blank page]

Introduction

8. Judging Investment Strategies

Reference:

Hakansson, Nils H., and William T. Ziemba (1995) "Capital Growth Theory," in: R.A. Jarrow, V. Maksimovic, and W.T. Ziemba, eds., *Handbooks in Operations Research and Management Science* 9: *Finance*, Amsterdam: Elsevier, 65-86.

Knight, Frank H. (1921) "The Meaning of Risk and Uncertainty," Part III, Chapter VII from *Risk, Uncertainty, and Profit*. Boston: Houghton and Mifflin. Online edition: http://www.econlib.org/library/Knight/knRUP.html.

Taleb, Nassim Nicholas (2001) *Fooled by Randomness: The Hidden Role of Chance in the Markets and in Life*. New York: Texere.

Track records of investors (mutual funds managers, etc.) are close to meaningless

Knightian uncertainty (Frank Knight, 1921)

> "It is the third type of probability or uncertainty which has been neglected in economic theory, … that higher form of uncertainty [is] not susceptible to measurement and hence to elimination. It is this *true uncertainty* [italicized in original] which … [prevents] … the theoretically perfect outworking …"
>
> In other words, we are not sure that the world in which we live is well charted

How could the hedge fund LTCM (Long-Term Capital Management) with its extraordinary talent (Nobel laureates Robert Merton and Myron Scholes of the Black-Scholes formula fame) blow up so spectacularly in 1998?

LTCM made no allowance for the possibility that they do not understand markets and that their methods are wrong

Put differently, LTCM fell prey to the Black Swan problem

The Scots philosopher David Hume posed this problem the following way (as phrased by John Stuart Mill):

> "No amount of observations of white swans can allow the inference that all swans are white, but the observation of a single black swan is sufficient to refute that conclusion."

In the summer of 1998, a black swan showed up on LTCM's computer screens when the Russian government defaulted on ruble-denominated debt and LTCM's hedging strategies failed

For details see the chapter "Professional Arbitrage (Limits of Arbitrage II)."

More generally, asset pricing models assume that at least one investor knows the true distribution of the asset's returns

For instance, when using a μ,σ-utility function, it is assumed that μ (the mean return) and σ (the standard deviation of the return) are known (to at least one investor)

Similarly, the Black-Scholes formula for option pricing assumes that the distribution of the returns of the underlying asset—in particular, its volatility σ—is known (to at least one investor).

Survivorship bias or "the problem of unwritten history"

- Many investors (mutual funds) have (to this point) survived because they were lucky (the lucky fools)
  - Yet, the public (as well as the fools) tends to attribute the survivors' wealth to a superior investment style, not noticing the carnage by the roadside.
- Imagine you drive through a suburb with 105 wonderful mansions
- Interested in the causes of these people's wealth, you ring the doorbells and ask about the owners' trades
- At each house, a 50-year old man answers the door and tells you that he has been playing Russian roulette since the age of 25
    - In Russian roulette, you put a revolver to your head with one of the six chambers loaded at random
    - Assume that the game is played once a year with a payoff (in the event of survival) of $1 million
    - If the initial cohort of 25-year olds numbered 10,000, we expect about 105 of them to be alive (and wealthy) at the age of 50
    - The problem is, there is also a big graveyard (of an expected 9,895 people), which goes unnoticed
      - The survivors enter the history books as "legendary investors" (rather than the lucky fools that they are), while the history of the dead remains unwritten.
- Thus, in judging investment styles it is important to add the losers to the set of observations
  - Unfortunately, winners are more visible than losers or, in other words, the highest performing realization is the most visible while the losers do not show up.

Trading rules (technical trading) and data snooping

Technical trading rules are developed by searching historical stock price data for mechanical trading rules that would have been profitable (back-testing)

Fitting the rule on the data is called data snooping

With sufficient computing power, it is very likely to detect trading rules that would have been profitable in the past even in a random data series

Note that random data series (data generated by stock prices that follow a random walk with drift) typically present some detectable pattern (for random data series rarely look random)

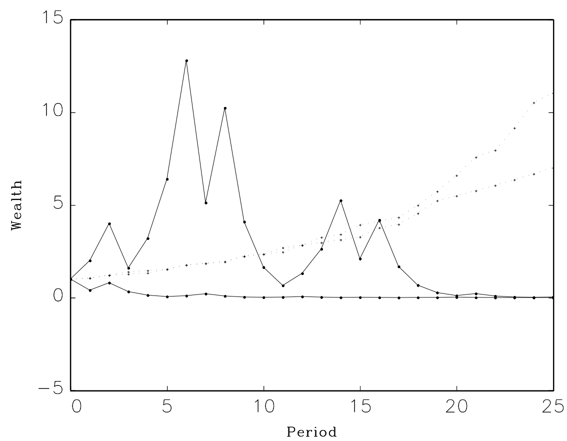
The problem with these trading rules is that, although they would have been profitable in the past, they have no predictive power for future asset returns.

Adverse selection

Investors (mutual funds) that advertise the most tend to be those that have been the most lucky, possibly because they took excessive risk

The figure below shows four return histories (out of an infinite number of possible histories) created by a Monte Carlo generator

Two investors invest an initial $1 into an asset that returns either 15 percent or 5 percent with equal probability (expected return: 10 percent; dotted line), while two other investors invest in an asset that returns either 100 percent or –60 percent with equal probability (expected return: 20 percent; solid line)



The risky mutual fund that “outperforms” the two prudent mutual funds advertises its performance in shiny investor magazines, pointing to its high cumulative five-year return

Naïve personal investors will buy into the fund and lose their money, as we will show below.

Note that neither of the two low-risk funds is superior to the other as judged by the generator (i.e., investment strategy)

Similarly, neither of the two high-risk funds is superior to the other.

There is a difference between maximizing expected return and survival in financial markets

The St. Petersburg Paradox—albeit not being a paradox—illustrates the difference between maximizing expected return and survival

The problem was studied by Daniel Bernoulli (1700-1782)

Imagine you are offered the following lottery

A coin is tossed until heads appears for the first time

The payoff, R, equals the number of tosses, n, before heads appears the first time:

$$R(n) = \$2^n$$

The probability of no tails ($n = 0$) equals $1/2$, the probability of one time tails and then heads ($n = 1$) equals $1/2 \times 1/2 \equiv (1/2)^2$, the probability of two times tails and then heads ($n = 2$) equals $1/2 \times 1/2 \times 1/2 \equiv (1/2)^3$, and so forth

The expected payoff is therefore:

$$\mathrm{E}(R) = \sum_{n=0}^{\infty} \mathrm{prob}(n) \times R(n) = \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^{n+1} \times 2^n = \sum_{n=0}^{\infty} \frac{1}{2} = \infty$$

Given that the payoff is infinite, should you pay all you can for a ticket that admits you to this lottery?

If your goal is to maximize expected final wealth, $E(w_T)$, your maximum willingness to pay (WTP) for participating in the lottery is unlimited; also, for a given price of the lottery ticket, you are willing to play this game an infinite number of times

It can easily be shown in a Monte Carlo study that if you bet your entire wealth on the lottery and you keep doing this forever, you end up poor with certainty:

$$\lim_{T \to \infty} w_T = \$1$$

In the limit, the probability that heads comes up on the first toss (in which event you are left with $1 of wealth) is unity.

The St. Petersburg Paradox highlights the important difference between survival in financial markets and maximizing expected final wealth (e.g., wealth at retirement)

If you keep going for the stocks with the highest expected returns, you keep going for the stocks with the highest risk; you will have little left to start over after a bad draw.

Logarithmic risk aversion

Consider the following family of utility functions:

$$U(w) = \frac{1}{\phi} w^{\phi} \ , \ \phi < 1$$

For $\phi = 0$, the logarithmic utility function results:

$$U(w) = \ln(w)$$

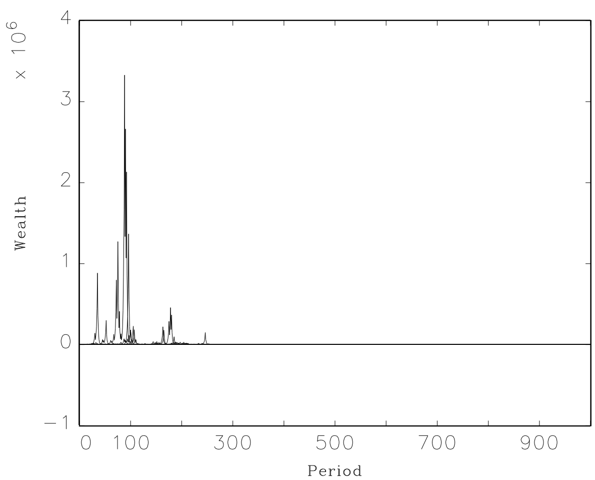
Note that logarithmic preferences have a relative risk aversion of unity:

$$r_r = -\frac{U''(w) \cdot w}{U'(w)} = -\frac{\frac{\partial^2 \ln(w)}{\partial w^2} \cdot w}{\frac{\partial \ln(w)}{\partial w}} = -\frac{-\frac{1}{w^2} \cdot w}{\frac{1}{w}} = 1$$

Remember that $\frac{\partial \ln(w)}{\partial w} = \frac{1}{w}$

It can be shown that an investment strategy that exhibits logarithmic risk aversion almost surely leads to more capital (in excess of what is offered by investing in the risk-free asset) in the long run than any other investment strategy that does not converge to it

- Let us return to the two investment strategies mentioned above
  - Invest an initial $1 in an asset that returns either 15 percent or 5 percent with equal probability (expected return: 10 percent)
  - Invest an initial $1 in an asset that returns either 100 percent or -60 percent with equal probability (expected return: 20 percent)
    - Note that this investment strategy—its expected return of 20 percent notwithstanding—implies almost sure ruin in the long run: the growth rate of capital converges to -10.56 percent!
      - The figure below plots the histories of 1,000 (out of an infinite number of possible) investors who follow the risky investment strategy over 1,000 periods; all of these investors face financial ruin (wealth below a penny), even though some of them at some point are millionaires, possibly celebrated in the business press and admired by the public as legendary investors



Back to the St. Petersburg Paradox

Maximizing expected final wealth by repeatedly risking your entire financial wealth in the Petersburg lottery leads to (almost sure) financial ruin in the long run

Which amount of money is an investor with logarithmic preferences willing to put at stake in this lottery?

With logarithmic preferences, the utility from playing the lottery equals

$$\mathrm{E}(U) = \mathrm{E}(\ln(w)) = \sum_{n=0}^{\infty} \mathrm{prob}(n) \times \ln(R(n))$$

$$= \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^{n+1} \times \ln(2^{n}) = \ln(2) \approx 0.693$$

Because $e^{0.693} = 2$, the investor's WTP for a lottery ticket is $\$2$.

Empirical Studies

9. Excessive Stock Market Volatility

Reference:

Shiller, Robert J. (1981) "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review* 71, 421-436.

> *"A conventional valuation which is established as the outcome of the mass psychology of a large number of ignorant individuals is liable to change violently as the result of a sudden fluctuation of opinion due to factors which do not really make much difference to the prospective yield; since there will be no strong roots of conviction to hold it steady. In abnormal times in particular, when the hypothesis of an indefinite continuance of the existing state of affairs is less plausible than usual even though there are no express grounds to anticipate a definite change, the market will be subject to waves of optimistic and pessimistic sentiment, which are unreasoning and yet in a sense legitimate where no solid basis exists for a reasonable calculation."*
>
> John Maynard Keynes
>
> *The General Theory of Employment, Interest and Money,* London: Harcourt Brace Jovanovich. 1964 (reprint of the 1936 edition), pp. 154.

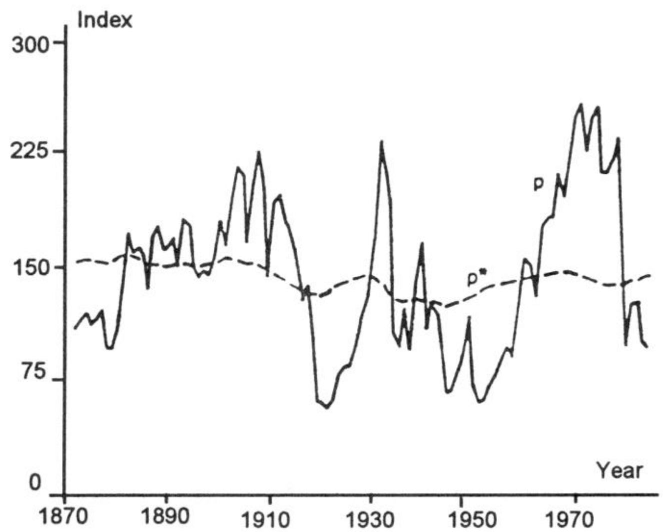
In a seminal article, Shiller (1991) argued that the stock market is more volatile than is justified by fundamentals

Shiller measures stock market fundamentals by the dividend stream the stock market generates

Shiller shows that stock market volatility is—depending on the metric—five to thirteen times too high to be attributed to new information about future real (i.e., inflation-adjusted) dividends if uncertainty about future dividends is measured by the standard deviations of real dividends around their long-run exponential growth path

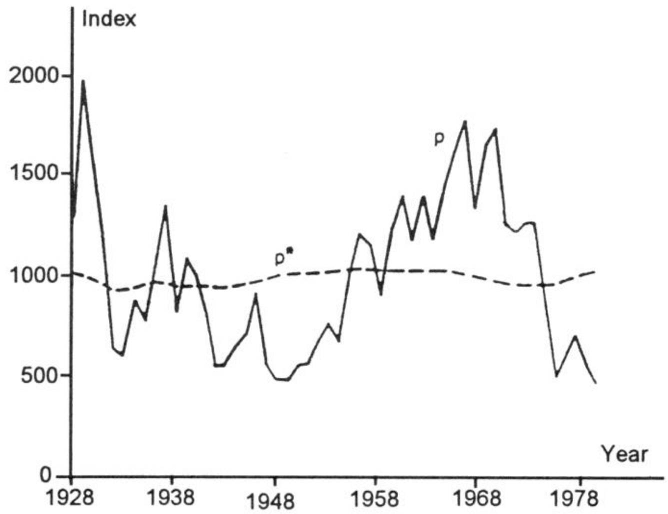
The following two figures illustrate the case for the S&P Composite Stock Price Index and the Dow Jones Industrial Average, respectively

Figure 1



Note: Real Standard and Poor's Composite Stock Price Index (solid line p) and *ex post* rational price (dotted line p^*), 1871–1979, both detrended by dividing a long-run exponential growth factor. The variable p^* is the present value of actual subsequent real detrended dividends, subject to an assumption about the present value in 1979 of dividends thereafter.

Figure 2



Note: Real modified Dow Jones Industrial Average (solid line p) and *ex post* rational price (dotted line p^*), 1928–1979, both detrended by dividing by a long-run exponential growth factor. The variable p^* is the present value of actual subsequent real detrended dividends, subject to an assumption about the present value in 1979 of dividends thereafter. Data are from Data Set 2, Appendix.

Source: Shiller (1981).

Shiller's theoretical argument runs as follows

The efficient markets hypothesis can be described, in its simplest form, as asserting that

$$p_t = E_t(p_t^*)$$

i.e., the stock price, p_t, is the mathematical expectation conditional on all information available at time t of the intrinsic value of the stock, p_t^*

In other words, p_t is the optimal forecast of p_t^*

Let $u_t = p_t^* - p_t$ be the forecasting error

A fundamental principle of unbiased forecasts is that the forecasting error, u_t, is uncorrelated with the forecast itself, p_t, that is the covariance between p_t and u_t must be zero

If the covariance between p_t and u_t is zero, we can write for the variance of u_t:

$$\text{var}(u_t) = \text{var}(p_t^*) - \text{var}(p_t)$$

Given that $\text{var}(u_t) \geq 0$, we can write:

$$\sigma_p \leq \sigma_{p^*}$$

In other words, the standard deviation of the forecast, σ_p, should be smaller than the standard deviation of the object of the forecast, σ_{p^*}

Remember that forecasts should gravitate toward the mean value, that is, be conservative: The less predictable a variable is, the closer the forecast should be to the historical mean (or historical growth path, if the mean is not stationary).

The figures above show that this inequality is grossly violated.

Shiller admits to two caveats, which have the potential of saving the general notion of efficient markets

First, the large swings in stock market valuation, p_t, compared to the ex-post rational counterpart, p_t^*, could be attributed to changes in the real interest rate (which determines the discount rate; see below)

Because expected real interest rates are not observable, such a hypothesis cannot be evaluated conclusively

However, the movements in expected real interest rates that would justify the variability in stock prices are very large—much larger than the movements in nominal interest rates over the sample period.

Second, the applied measure of uncertainty regarding future dividends—the sample standard deviation of the movements of real dividends around their long-run exponential growth path—might understate the true uncertainty about future dividends

Perhaps the market was rightfully fearful of much larger movements than actually materialized

Remember that the fact that catastrophic events did not occur does not imply that the market was wrong in considering them.

Table 2 Sample Statistics for Price and Dividend Series

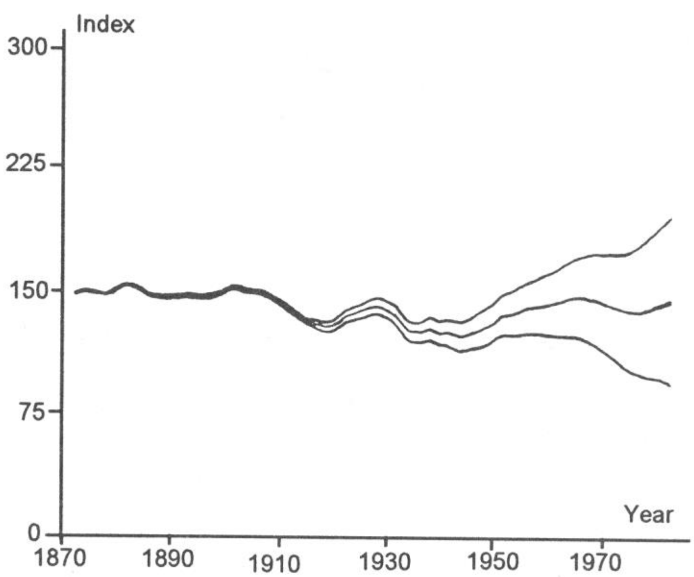
	Data Set 1: Standard and Poor's	Data Set 2: Modified Dow Industrial
Sample Period:	1871–1979	1928–1979
1) $E(p)$	145.5	982.6
$E(d)$	6.989	44.76
2) $\bar{r}$	.0480	0.456
$\bar{r}_2$	.0984	.0932
3) $b = \ln \lambda$	.0148	.0188
$\hat{\sigma}(b)$	(.0011)	(1.0035)
4) $\text{cor}(p, p^*)$	.3918	.1626
$\sigma(d)$	1.481	9.828
Elements of Inequalities:		
Inequality (1)		
5) $\sigma(p)$	50.12	355.9
6) $\sigma(p^*)$	8.968	26.80
Inequality (11)		
7) $\sigma(\Delta p + d_{-1} - \bar{r}p_{-1})$	25.57	242.1
$\min(\sigma)$	23.01	209.0
8) $\sigma(d)/\sqrt{\bar{r}_2}$	4.721	32.20
Inequality (13)		
9) $\sigma(\Delta p)$	25.24	239.5
$\min(\sigma)$	22.71	206.4
10) $\sigma(d)/\sqrt{2\bar{r}}$	4.777	32.56

Note: In this table, E denotes sample mean, σ denotes standard deviation and $\hat{\sigma}$ denotes standard error. Min(σ) is the lower bound on σ computed as a one-sided χ^2 95 percent confidence interval. The symbols p, d, $\bar{r}$, $\bar{r}_2$, b, and p^* are defined in the text. Data sets are described in the Appendix. Inequality (1) in the text asserts that the standard deviation in row 5 should be less than or equal to that in row 6, inequality (11) that σ in row 7 should be less than or equal to that in row 8, and inequality (13) that σ in row 9 should be less than that in row 10.

Table 1 Definitions of Principal Symbols

γ =	real discount factor for series before detrending; $\gamma = 1/(1 + r)$
$\bar{\gamma}$ =	real discount factor for detrended series; $\bar{\gamma} \equiv \lambda\gamma$
D_t =	real dividend accruing to stock index (before detrending)
d_t =	real detrended dividend; $d_t \equiv D_t/\lambda^{t+1-T}$
Δ =	first difference operator $\Delta x_t \equiv x_t - x_{t-1}$
δ_t =	innovation operator; $\delta_t x_{t+k} \equiv E_t x_{t+k} - E_{t-1} x_{t+k}$; $\delta x \equiv \delta_t x_t$
E =	unconditional mathematical expectations operator. $E(x)$ is the true (population) mean of x
E_t =	mathematical expectations operator conditional on information at time t; $E_t x_t \equiv E(x_t \mid I_t)$ where I_t is the vector of information variables known at time t
λ =	trend factor for price and dividend series; $\lambda \equiv 1 + g$ where g is the long-run growth rate of price and dividends
P_t =	real stock price index (before detrending)
p_t =	real detrended stock price index; $p_t = P_t/\lambda^{t-T}$
p_t^* =	*ex post* rational stock price index (expression 4)
r =	one-period real discount rate for series before detrending
$\bar{r}$ =	real discount rate for detrended series; $\bar{r} = (1 - \bar{\gamma})/\bar{\gamma}$
$\bar{r}_2$ =	two-period real discount rate for detrended series; $\bar{r}_2 = (1 + \bar{r})^2 - 1$
t =	time (year)
T =	base year for detrending and for wholesale price index; $p_T = P_T$ = nominal stock price index at time T

Figure 3



Note: Alternative measures of the *ex post* rational price p^*, obtained by alternative assumptions about the present value in 1979 of dividends thereafter. The middle curve is the p^* series plotted in Figure 1. The series are computed recursively from terminal conditions using dividend series d of Data Set 1.

Dividends versus earnings

According to the simple efficient-markets model employed by Shiller, the real (i.e., inflation-adjusted) price, P_t, of a share at the beginning of the time period t is given by

$$P_t = \sum_{k=0}^{\infty} \frac{\mathrm{E}_t\,(D_{t+k})}{(1+r)^{k+1}}\ , \qquad 0 < \frac{1}{1+r} < 1$$

where D_t is the real dividend paid the end of period t, E_t denotes the mathematical expectation conditional on information available at time t, and r is the (constant) real risk-free rate.

It has been argued that Shiller's model should be replaced by a model which makes price the present value of expected (real) earnings (rather than (real) dividends)

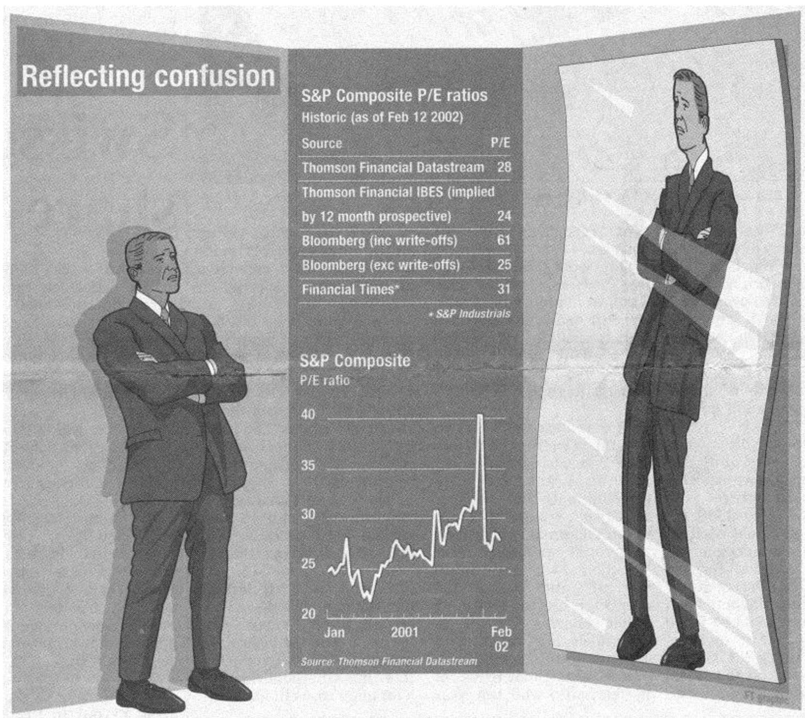
In Shiller's model, on the other hand, earnings may be relevant to the pricing of shares but only insofar as earnings are indicators of future dividends

Earnings are thus no different from any other economic variable that may indicate future dividends (and thus are part of the information set captured by E_t).

The case against earnings

First, Shiller's model implies that expected total returns are constant and that the capital gains component of returns is just a reflection of information about future dividends

Earnings, in contrast, are statistics conceived by accountants that are supposed to provide an indicator of how well a company is doing, and there is a great deal of latitude for the definition of earnings



Source: Financial Times, Thursday February 14, 2002.

Note the difference in the two Bloomberg numbers

According to Zacks Investment Research, 247 of the S&P 500 companies reported negative non-recurring items in 2000.

Note that the Datastream number corresponds to the S&P number used in the chapter “Stock Market Valuation.”

Second, there is no reason why price per share ought to be the present value of expected earnings per share if some earnings are retained (rather than paid out as dividends)

In fact, as Merton H. Miller and Franco Modigliani (1961; "Dividend Policy, Growth and the Valuation of Shares," *Journal of Business* 34, 411-33) argued, such a present value formula wold entail double counting

It is incorrect to include in the present value both earnings at time *t* and the later earnings that accrue when time *t* earnings are reinvested

Miller and Modigliani presented a formula by which price might be regarded as the present value of earning corrected for investment, but that formula can be shown to be identical to Shiller's dividend-based valuation model.

Empirical Studies

10. The Closed-End Fund Puzzle

Reference:

Lee, Charles M.C., Andrei Shleifer, and Richard H. Thaler (1991) "Investor Sentiment and the Closed-End Fund Puzzle," *Journal of Finance* 46, 75-109.

Shleifer, Andrei (2000) *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford: Oxford University Press, Chap. 3.

Note: The figures and tables in this chapter, if not indicated otherwise, are from Lee, Shleifer, and Thaler (1991)

The closed-end fund puzzle is one of the most perplexing puzzles in finance

A closed-end fund, like the more popular open-end fund, is a mutual fund that typically holds other publicly traded securities

Unlike an open-end fund, a closed-end fund issues a fixed number of shares that are traded on the stock market

To liquidate a holding in a closed-end fund, investors must sell their shares to other investors rather than redeem them with the fund itself for the net asset value (NAV) per share as they would with an open-end fund.

The closed-end fund puzzle is the empirical finding that closed-end fund shares typically sell at prices not equal to the share market value of assets the fund holds

Closed-end funds tend to trade at discounts of often 10 or 20 percent, after selling at premiums in the initial public offering (IPO).

Finance scholars have attempted to solve the closed-end fund puzzle through various avenues

- Agency costs
  - The agency costs hypothesis tries to explain the discount by the funds' operating expenses and by potential sub-par managerial performance.
- Tax liabilities
  - The tax explanation argues that the discount is caused by capital gains tax liabilities on unrealized appreciation (at the fund level), which are not captured by the standard calculation of NAV.
- Illiquid assets
  - The argument has been made that closed-end funds hold (to some extent) illiquid assets, which are overvalued in the calculation of NAV (because they may not be tradable at the quoted price)
  - While there is some empirical support for this hypothesis, it should be noted that many open-end funds, in particular large ones, hold only very small fractions of their assets in illiquid securities.

In conclusion, it is fair to say that none of the three suggested explanations, separately or together, explain the closed-end fund puzzle in all its dimensions, which are to be discussed below.

Lee, Shleifer, and Thaler (henceforth "the authors") try to explain the closed-end fund puzzle (in all its dimensions) by what they call investor sentiment

There are four important dimensions of the closed-end fund puzzle, which together characterize the life-cycle of the closed-end fund

First, closed-end funds start out at a premium of almost 10 percent (above NAV), when organizers raise money from new investors and use it to purchase securities

Most of this premium results from underwriting and start-up costs

The reason that investors pay a premium for new funds when old funds trade at a discount is to be explained.

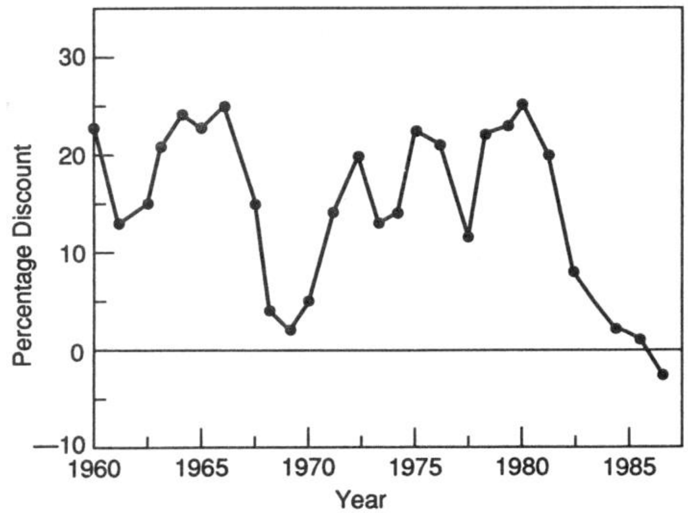
Second, although they start at a premium, closed-end funds move to an average discount of over 10 percent within 120 days from the beginning of trading, and keep trading at a discount henceforth

For illustration, the following figure shows the year-end discounts on the Tricontinental Corporation (TRICON) fund during 1960-1986

TRICON is the largest closed-end fund trading on U.S. exchanges, with net assets of over $1.3 billion as of December 1986

For most of the time during the 27 year period, the fund trades at a discount, which frequently hovers around 20 percent

Figure 1 Percentage discount or premium of Tricontinental Corporation at the end of each year during 1960–1986.



The percentage discount is computed as 100 × (NAV − SP); where NAV is the per share net asset value and SP is the share price of the fund. The mean (median) of the percentage discount or premium is 14.43 (15.0). The maximum (minimum) value is 25.0 (−2.5) and the standard deviation is 8.56.

Third, as the figure above illustrates, discounts on closed-end funds are subject to wild fluctuations over time

Empirical studies document mean reversion in closed-end funds discounts and abnormal returns from assuming long positions on funds with large discounts.

Fourth, when closed-end funds are terminated through either a liquidation or an open-ending, share prices rise and discounts shrink

Most of the positive returns to shareholders accrue when discounts narrow around the announcement of termination; a small discount persists, however, until final termination or open-ending.

Investor sentiment

The authors make an attempt to explain the four dimensions of the closed-end fund puzzle with investor sentiment

The concept of investor sentiment is closely related to the concept of noise traders

Noise traders, unlike sophisticated investors, are unable to distinguish information from noise

Noise traders might show signs of excessive optimism or pessimism that manifests itself in investor sentiment

Sophisticated investors issue closed-end funds to unsophisticated investors when the latter are particularly optimistic about stock market returns

Without such a snake oil salesman approach, it is hard to explain why investors pay a premium for a closed-end fund at the IPO, given that shortly thereafter the fund is likely to trade at a discount (and will continue to do so).

Noise traders are a major source of asset mispricing, because they may cause sustained deviations of prices from the securities' intrinsic values

Noise traders add to the risk of securities if the time path of investor sentiment is not predictable

Arbitrageurs who try to eliminate asset mispricing by going long on (relatively) underpriced assets and, simultaneously, short on relatively overpriced assets, face the risk that the mispricing deepens temporarily

If the time horizon of arbitrageurs were infinite, temporary deepening of mispricing would not matter.

To the degree that investor sentiment is correlated across noise traders, noise trader risk is systematic (rather than idiosyncratic), which means that it cannot be eliminated through diversification and is consequently priced

In other words, securities that are affected by noise trader risk, trade at a discount relative to securities with identical cash flows that are not affected by noise trader risk

The discount is caused by the higher expected return investors demand for the higher systematic risk

Translated into the world of closed-end funds, if the closed-end fund is affected by noise trader risk while the securities the fund holds are not, the fund is more risky than its portfolio, and consequently the fund trades at a discount (after trading at a premium at and shortly after the IPO)

Note that the discount can be explained without assuming pessimistic investor sentiment

Remember that the figure above is representative in that it shows a premium at the IPO and shortly thereafter, followed by a persistent (and fluctuating) discount.

Why is the fund more affected by noise trader risk than its portfolio?

Closed-end funds are owned predominantly by small investors (and thus noise traders) while this category of investors is less strongly represented in the ownership of the portfolios these funds hold

The authors found the average institutional ownership in the closed-end mutual funds in their sample at the beginning of 1988 to be just 6.6 percent (median 6.2 percent)

By comparison, average institutional ownership for a random sample of the smallest 10 percent of NYSE (New York Stock Exchange) stocks is 26.5 percent (median 23.9 percent), and 52.1 percent (median 54.0 percent) for the largest 10 percent of NYSE stocks

Note that the smaller fraction of institutional ownership in small caps (small capitalization stocks) reported by the authors is no aberration; although the growth of small cap mutual funds has increased, institutional ownership in small companies in the 1990s, it is still much lower than in large companies.

Why don't arbitrageurs eliminate the mispricing (the closed-end fund discount) by going long on closed-end funds and short on their portfolios?

First, there is noise trader risk, which we will discuss in the chapter "Noise Trader Risk (Limits of Arbitrage I)"

Second, the holdings of mutual funds are published at the end of the quarter only, which means that an arbitrageur cannot perfectly track the funds' portfolios

Arbitrage becomes a powerful force once the fund announces an open-ending or liquidation (and the noise trader risk goes away)

> As mentioned, upon announcement of open-ending or liquidation, most of the discount goes away.

Data and variable description for the basic analysis

The authors started out with a total of 87 funds, of which 68 were selected for monthly analysis because they were known to have CUSIP identifiers (http://www.cusip.com)

For these funds, the authors collected the weekly net asset value (NAV) per share, stock price, and discount per share as reported by the *Wall Street Journal* (WSJ) <u>from July 1956 through December 1985</u>

The NAV-per-share information was then combined with the number of shares outstanding at the end of each month (as obtained from the monthly CRISP [Center for Research of Security Prices at the University of Chicago] master tape to arrive at the total net asset value for each fund

For several of the tests the authors constructed, a value-weighted index of discounts (VWD) both at the annual and monthly levels as follows:

$$VWD_t = \sum_{i=1}^{n_t} W_i \cdot \mathrm{DISC}_{it}$$

where $$W_i = \frac{\mathrm{NAV}_{it}}{\sum_{i=1}^{n_t} \mathrm{NAV}_{it}} ,$$

NAV_{it} = net asset value of fund i at end of period t

$$\mathrm{DISC}_{it} = \frac{\mathrm{NAV}_{it} - \mathrm{SP}_{it}}{\mathrm{NAV}_{it}} \times 100$$

SP_{it} = stock price of fund i at end of period t

n_t = number of funds with available DISC_{it} and NAV_{it} data at the end of period t

The authors also computed changes in the value-weighted index of discounts

$$\Delta VWD_t = VWD_t - VWD_{t-1},$$

which can be done only for funds with available DISC_{it} and NAV_{it} data at the end of periods t and $t-1$

Of the original 68 funds, 18 were either missing data from the WSJ or did not have shares information available on CRSP, and 30 others were bond funds

This left a total of 20 stock funds which participated in the monthly ΔVWD series

Of these remaining funds, some had relatively short life spans, others may occasionally have missing data, so the actual number of funds included in computing VWD and ΔVWD varied from month to month

In the vast majority of months, at least 10 funds were in the index.

Empirical evidence

The concept of investor sentiment which manifests itself primarily in investment decisions of personal investors—who constitute the bulk of noise traders—suggests that …

… the discounts are correlated across funds (cross-sectional <u>co-movement of discounts</u>)

… the discounts are largely uncorrelated with stock-market returns (for personal investors are less strongly represented among owners of stocks of individual corporations than they are among closed-end funds shareholders).

The following figure shows the levels of discounts for the entire sample of closed-end funds at the end of each year during 1960-1986

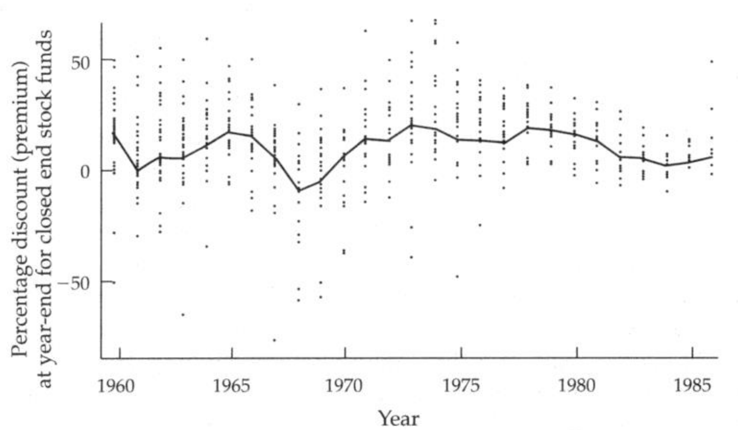


FIG. 3.2. Percentage discount or premium at the end of the year for all closed end stock funds during 1960-86.
The percentage discount is computed as 100 × (NAV − SP)/NAV; where NAV is the per share net asset value and SP is the share price of the fund. The sample includes all 46 stock funds reported in the Wiesenberger Investment Companies Services Annual survey during this period. The discount on a value-weighted portfolio of these funds is represented by the solid line.

Source: Shleifer (2000, p. 68).

When do get funds started?

The investor sentiment to pricing of closed-end funds suggests that new funds get started when existing funds sell at premiums or at small discounts

The testing of this hypothesis poses several problems

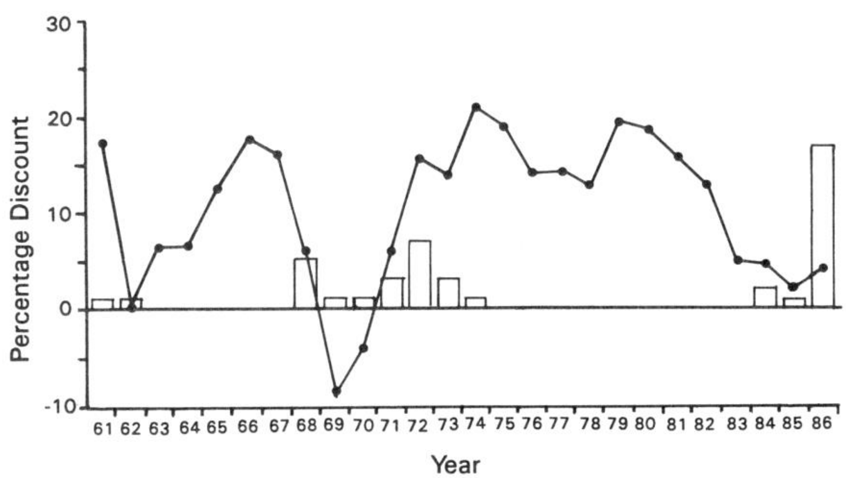
First, over most of the time period that the authors examine, very few funds get started

Second, it takes time to organize and register a fund, which means that funds may start trading much later than the time at which they are conceived, or might even be withdrawn.

The following figure plots the number of funds started each year against the *VWD* (value-weighted index of discounts) at the beginning of the year

The fund starts tend to be clustered and coincide with periods when discounts are relatively low

Figure 3 The number of closed-end stock funds started and the discount on stock funds at the beginning of the year.



This graph shows the number of closed-end stock funds started during the year and the percentage discount on a value-weighted portfolio of closed-end stock funds at the beginning of each year during 1961 to 1986. The line graph represents the percentage discount at the beginning of the year. The bar graph represents the number of stock funds started during the year.

The following table (authors' Table III) compares the value-weighted discounts on seasoned funds in years when one or more new stock funds begin trading with the corresponding discounts in years where no stock funds begin trading

- From 1961 through 1986, there are 12 years in which one or more stock funds get started and 14 years in which no stock funds is established
  - The average beginning-of-year discount in the former years is 6.40 percent, compared with 13.64 percent in the latter years; the difference is statistically significant

Table III Statistical Comparison of the Value-Weighted Discount at the Beginning of the Year for Years with Fund Starts and Years without Fund Starts

Statistical comparison of the value-weighted discount at the beginning of the year for years in which *one or more closed-end stock funds* were started versus the years in which no stock funds started.**

	Years in which one or More Stock Funds Started	Years in which No Stock Funds Started
Mean value-weighted discount at the beginning of the year	6.40	13.64
Number of years	12	14
t-statistic for a test of a difference in means between two random samples assuming unequal variance	−2.51**	
t-statistic for a test of a difference in means between two random samples assuming equal variance	−2.63**	
z-statistic for the Wilcoxon rank sum test of a difference in means between two random samples	−2.24**	

**Significant at the 1% level in one-tailed tests (5% in two-tailed tests).

Discount movements and returns on portfolios of stocks

- As mentioned, individual investors are significant holders and traders of small-cap stocks but less so of large-cap stocks
  - The concept of investor sentiment suggests that the change in the value-weighted index of discounts, ΔVWD, correlates strongly with returns on small caps but less so—or not at all— with return on large caps or the market overall
    - The market portfolio is the value-weighted portfolio of companies listed at the New York Stock Exchange, VWNY
    - The authors construct ten size-rank portfolios of NYSE-listed stocks.

Column "ΔVWD" of the following table (authors' Table IV) supports the hypothesis that the discount change correlates more strongly with small-cap than large-cap portfolios

The (absolute value of the) regression coefficient in column "ΔVWD" is monotonically decreasing with size decile, while for the largest decile the regression coefficient is even negative

Table IV The Time-Series Relationship between Returns on Size-Decile Portfolios, the Market Return, and Changes in Closed-End Fund Discounts

The time-series relationship (7/65 to 12/85) between monthly returns on decile portfolios (dependent variables), changes in the monthly discount on a value-weighted portfolio of closed-end stock funds (ΔVWD), and the monthly return on a value-weighted portfolio of New York Stock Exchange firms (VWNY). Decile 10 contains the largest firms, Decile 1 the smallest. Membership in each decile is determined at the beginning of year and kept constant for the rest of the year. Returns of each firm are weighted by the beginning-of-month market capitalization. In case of missing returns, a firm is excluded from the portfolio for the current and following month. The dependent variable in the last row is the excess return of small firms over large firms, computed by subtracting Decile 10 returns from Decile 1 returns. The number of observations is 245. *t*-statistics are shown in parentheses.

Return on the Decile Portfolio	Intercept	ΔVWD	VWNY	Adjusted R^2
1	0.0062	−0.0067	1.238	58.7
(smallest)		(−4.94)	(18.06)	
2	0.0042	−0.0049	1.217	70.3
		(−4.83)	(23.66)	
3	0.0036	−0.0039	1.202	74.0
		(−4.20)	(26.09)	
4	0.0033	−0.0038	1.163	79.7
		(−5.07)	(30.64)	
5	0.0027	−0.0029	1.148	81.8
		(−4.12)	(32.90)	
6	0.0024	−0.0028	1.124	85.1
		(−4.65)	(37.08)	
7	0.0013	−0.0015	1.134	89.4
		(−3.03)	(45.30)	
8	0.0015	−0.0015	1.088	91.5
		(−3.45)	(51.32)	
9	0.0003	−0.0010	1.057	94.8
		(−3.14)	(66.93)	
10	−0.0005	0.0010	0.919	95.4
(largest)		(3.84)	(71.34)	
1–10	0.0067	−0.0077	0.319	13.5
		(−4.93)	(4.05)	

The results from the authors' Table IV (above) hold up when the sample period is split into two sub-samples (see authors' Table VII below)

- The motivation for splitting the sample is the steady <u>increase in institutional ownership in small companies</u> over time
  - As mentioned, 26.5 percent of the shares of the smallest-decile companies were held by institutions in 1988
  - An examination of a random sample of the smallest decile-companies in 1980 revealed that institutions held only 8.5 percent of the shares, which means that in just 8 years, institutions more than tripled their holding in first-decile companies.
- Consistent with the increased importance of institutions as small-cap shareholders, the results of Table IV are not as strong for the second half of the sample period (10/1975-12/1985) than the first (7/1967-9/1975), as shown in the table below (authors' Table VII).

In summary, the evidence suggests that discounts on closed-end funds narrow when small-cap stocks do well

- This is consistent with the discount narrowing as individual investors turn optimistic and widening as they turn pessimistic.

Table VII Stability of the Time-Series Relationship between Returns on Size-Decile Portfolios, the Market Return, and Changes in Closed-End Fund Discounts

Analysis of the stability of the time-series relationship between monthly returns on decile portfolios (dependent variables), changes in the monthly discount on a value-weighted portfolio of closed-end stock funds (ΔVWD) and the monthly return on a value-weighted portfolio of New York Exchange firms (VWNY). Decile 10 contains the largest firms, Decile 1, the smallest. Membership in each decile is determined at the beginning of year and kept constant for the rest of the year. Returns of each firm is weighted by the beginning-of-month market capitalization. In case of missing returns, a firm is excluded from the portfolio for the current and following month. The dependent variable in the last row is the excess return of small firms over large firms, computed by subtracting Decile 10 returns from Decile 1 returns. The number of observations for the first period is 122, the second period is 123. *t*-statistics are shown in parentheses.

Return on the Decile Portfolio	First 123 months (7/65 to 9/75)				Second 123 months (10/75 to 12/85)			
	Intercept	ΔVWD	VWNY	Adj. R^2	Intercept	ΔVWD	VWNY	Adj. R^2
1 (smallest)	0.0054	−0.0101 (−5.50)	1.355 (13.83)	63.2	0.0079	−0.0022 (−1.08)	1.140 (12.08)	54.9
2	0.0015	−0.0070 (−4.89)	1.303 (16.97)	71.1	0.0078	−0.0022 (−1.52)	1.129 (16.79)	70.3
3	0.0016	−0.0057 (−4.60)	1.269 (19.18)	75.6	0.0064	−0.0014 (−1.00)	1.137 (17.80)	72.5
4	0.0022	−0.0050 (−4.88)	1.206 (21.99)	80.2	0.0048	−0.0022 (−1.98)	1.123 (21.16)	79.1
5	0.0010	−0.0042 (−4.59)	1.193 (24.27)	83.1	0.0050	−0.0010 (−0.95)	1.104 (22.29)	80.5
6	0.0014	−0.0038 (−4.58)	1.184 (26.79)	85.6	0.0041	−0.0016 (−1.81)	1.060 (25.71)	84.7
7	0.0006	−0.0021 (−2.90)	1.184 (31.04)	88.8	0.0025	−0.0009 (−1.31)	1.080 (33.44)	90.3
8	0.0016	−0.0018 (−2.98)	1.123 (35.67)	91.3	0.0017	−0.0012 (−1.89)	1.053 (36.56)	91.8
9	0.0000	−0.0013 (−2.82)	1.084 (44.58)	94.3	0.0009	−0.0007 (−1.52)	1.027 (50.93)	95.6
10 (largest)	−0.0002	0.0014 (4.16)	0.902 (50.18)	95.5	−0.0010	0.0004 (1.11)	0.937 (50.12)	95.4
1–10	0.0056	−0.0115 (−5.47)	0.4530 (4.04)	25.2	0.0089	−0.0027 (−1.12)	0.2038 (1.87)	2.5

Do closed-end funds hold small-cap stocks?

Small-cap stocks might trade infrequently, which means that reported prices are stale, causing mismeasurement of the fund's NAV

On the other hand, the fund itself might be frequently traded and prices thus be relatively fresh

Taken together, the measured discount might exaggerate the difference between the market capitalization of the fund and the actual NAV during times when small caps are thinly traded (and prices update infrequently)

When a <u>change in beliefs</u> sets in, for instance, when investors turn optimistic, turnover increases and prices update more frequently, narrowing the discount

The measured correlation between returns on small caps and the discount would thus be an artifact of stale prices of the shares the funds hold.

The following table (authors' Table V) studies the portfolio holding of the aforementioned TRICON fund

The table shows the distribution of the holdings by size decile of the stock's market capitalization, every five years, starting in 1965

The holdings are concentrated in stocks of the largest two deciles, which, together with short-term holdings and cash equivalents, represent abut 80 percent of the fund's holdings

In contrast, the fund typically holds less than 4 percent of its assets in stocks that belong to the five lowest deciles.

Table V Composition of the Tricontinental Corporation Investment Portfolio

Composition of the investment portfolio of Tricontinental Corporation (Tricon) at the end of the year, distributed by the total market capitalization of the individual investments. To construct this table, each holding in the Tricon portfolio for each of the years listed was identified from the financial statements of the fund. For the majority of holdings, market capitalization was obtained through the CRSP tapes; market capitalization for the remainder were traced to Moody's Security Manuals and manually checked against Decile cutoffs for each year. Values are shown in thousands of dollars. Decile cutoffs for each year are the same as those used on earlier regressions and are obtained from CRSP. Cash and short-term holdings include government T-bills and corporate debt instruments, net of short-term liabilities of the fund. Other holdings represent equity securities for which the market capitalization was not readily obtainable.

	1985		1980		1975		1970		1965	
Decile 1	0.0	0.0	0.0	0.0	2902.4	0.5	3644.7	0.6	8486.8	1.5
Decile 2	0.0	0.0	3316.5	0.4	548.5	0.1	7514.0	1.2	5856.0	1.0
Decile 3	2793.8	0.2	0.0	0.0	3507.9	0.6	125.8	0.0	0.0	0.0
Decile 4	0.0	0.0	7000.0	0.8	2051.2	0.4	1575.0	0.3	0.0	0.0
Decile 5	2477.9	0.2	19125.0	2.2	9840.5	1.7	9715.5	1.6	8016.2	1.4
Decile 6	4575.0	0.4	38519.2	4.4	5903.5	1.0	14304.3	2.4	0.0	0.0
Decile 7	63575.5	5.3	58238.9	6.6	28283.5	5.0	21934.8	3.7	23832.0	4.3
Decile 8	118981.2	10.0	88204.4	10.1	53320.2	9.4	51241.0	8.5	76452.2	13.7
Decile 9	306874.7	25.7	181298.3	20.7	69407.0	12.2	49787.5	8.3	82263.8	14.7
Decile 10	558993.8	46.8	391753.9	44.7	344500.4	60.7	371398.4	61.7	336612.2	60.2
Short-term holdings & cash equivalents	128745.1	10.8	67978.2	7.8	41905.7	7.4	60690.5	10.1	17940.0	3.2
Other holdings	8143.2	0.7	20890.9	2.3	5474.4	1.0	9702.1	1.6	0.0	0.0
Total value of portfolio	1195160.3	100.0%	876325.3	100.0%	567645.2	100.0%	601633.6	100.0%	559459.2	100.0%

The following table (authors' Table VI) repeats the exercise from the authors' Table IV, but this time for TRICON only

That is to say, the authors regress the return on size-ranked portfolio deciles on the VWNY (value-weighted NYSE portfolio) return and the change in the discount of TRICON (rather than the change in the value-weighted discount, ΔVWD)

The table shows that TRICON's discount narrows when small caps do well, <u>although TRICON holds virtually no small caps</u>

Table VI The Time-Series Relationship between Returns on Size-Decile Portfolios, the Market Return, and Changes in the Discount of Tri-Continental Corporation.

The time-series relationship (7/65 to 12/85) between monthly returns on decile portfolios (dependent variables), changes in the monthly discount of Tri-Continental (TriCon) and the monthly return on a value-weighted portfolio of New York Stock Exchange firms (VWNY). Decile 10 contains the largest firms, Decile 1 the smallest. Membership in each decile is determined at the beginning of year and kept constant for the rest of the year. Returns of each firm is weighted by the beginning-of-month market capitalization. In case of missing returns, a firm is excluded from the portfolio for the current and following month. The dependent variable in the last row is the excess return of small firms over large firms, computed by subtracting Decile 10 returns from Decile 1 returns. The number of observations is 241. *t*-statistics are shown in parentheses.

Return on the Decile Portfolio	Intercept	TriCon	VWNY	Adjusted R^2
1	0.0062	−0.0026	1.263	56.0
(smallest)		(−2.74)	(17.52)	
2	0.0044	−0.0021	1.236	68.9
		(−2.98)	(23.11)	
3	0.0039	−0.0017	1.214	72.9
		(−2.70)	(25.46)	
4	0.0036	−0.0013	1.174	78.3
		(−2.41)	(29.39)	
5	0.0030	−0.0011	1.156	81.0
		(−2.40)	(31.96)	
6	0.0025	−0.0014	1.135	84.6
		(−3.41)	(36.28)	
7	0.0014	−0.0009	1.142	89.4
		(−2.76)	(44.99)	
8	0.0016	−0.0010	1.097	91.7
		(−3.54)	(51.41)	
9	0.0004	−0.0007	1.062	94.8
		(−3.21)	(66.21)	
10	−0.0006	0.0005	0.916	95.4
(largest)		(2.94)	(69.80)	
1–10	0.0069	−0.0031	0.347	8.1
		(−2.85)	(4.20)	

Evidence from open-end fund redemption

The authors regress redemption on open-end mutual funds on the change in the value-weighted index of discounts, ΔVWD, and the return on the market portfolio, VWNY (value-weighted portfolio of companies listed at the New York Stock Exchange)

The authors employ two concepts of redemption ratios

"R/S", which is the ratio of net redemption to sales

"NRED", which is the ratio of "net redemption minus sales" to the fund's assets.

If changes in closed-end discounts are driven by changes in investor sentiment, (after controlling for changes in market returns) changes in closed-end discount can be expected to vary with open-end fund redemption

The following table (authors' Table XI) confirms this hypothesis

Note that the authors cut off the post-2/1985 period in Panel B, because of the subsequent enormous increase in net purchases of open-end funds

Also, note that the authors did not use the dependent variable "R/S" in logarithmic form (as can be judged by the value of the intercept)

The ratio "R/S" is log-normally (rather than normally) distributed, because zero is a lower bound.

Table XI The Relationship Between Net Redemption on Open-End Funds, the Market Return, and Changes in the Value-Weighted Discount

The time-series relationship between net redemption on open-end funds (dependent variable), the monthly return on a value-weighted portfolio of New York Stock Exchange firms (VWNY), and changes in the monthly discount on a value-weighted portfolio of closed-end stock funds (ΔVWD). The net redemption on open-end funds is measured two ways: by the monthly ratio of net redemptions to sales on open-end funds (R/S) and by the monthly net redemption on open-end funds expressed as a percentage of total funds assets at the beginning of the month (NRED). R/S is computed as redemptions/sales. NRED is computed as (redemptions-sales)/total fund assets. Monthly redemptions, sales, and fund assets data are obtained from the Investment Companies Institute and represent all open-end funds with long-term investment objectives (i.e., exclude money market and short-term municipal bond funds). *t*-statistics are shown in parentheses.

Model	Dep. Var.	Intercept	VWNY	ΔVWD	Adj. R^2	No. of Obs.
PANEL A—7/65 to 12/85						
1	R/S	0.855	−1.864	0.029	4.9	245
			(−3.03)	(2.35)		
2	NRED	−0.005	−0.044	0.0001	3.0	245
			(−3.05)	(0.38)		
PANEL B—7/65 to 2/82						
1	R/S	0.949	−1.417	0.034	4.5	199
			(−2.18)	(2.53)		
2	NRED	−0.001	−0.009	0.0003	3.6	199
			(−1.73)	(2.50)		

Evidence from Initial Public Offerings (IPOs)

Another domain in which individual investors are important is initial public offerings (of companies other than closed-end funds)

The investor sentiment hypothesis suggests that these IPOs should be more prevalent in times when individual investors are optimistic, i.e., when stocks fetch high prices relative to their fundamental values

See the chapter "Long-Run Performance of IPOs"

While institutions are more important buyers of IPOs than they are of closed-end funds, individuals still account for more than 75 percent of IPO subscriptions (at the time this study was written).

The authors regress the annual number of IPOs on the value-weighted index of discounts, VWD, and the dividend yield of the S&P 500 Stock Price Index, a measure of which the authors assume that it captures overall market conditions

Note that the number of IPOs might be log-normally (rather than normally) distributed; consequently, the authors should have applied a logarithmic transformation

The authors' Table XII supports the hypothesis that IPO activity and closed-end fund discount are highly correlated, possibly driven by the same underlying force—investor sentiment

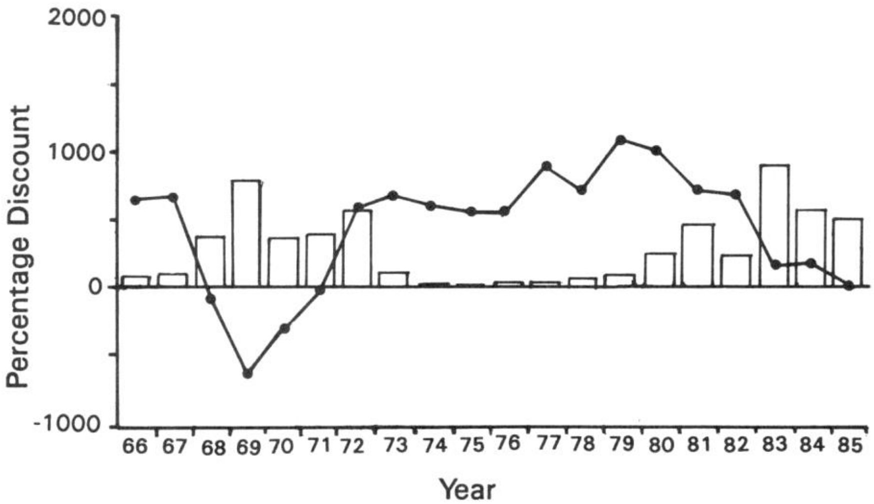
Table XII The Relationship between Number of IPO's, the Dividend-to-Price Ratio on S&P500, and the Value-Weighted Discount at the Beginning of the Year

The time-series relationship between the annual number of Initial Public Offerings (dependent variable), the dividend to price ratio of S&P500 stocks at the beginning of the year expressed as a percentage (Div/Price), and the level of the value-weighted discount on a portfolio of closed-end funds at the beginning of the year (VWD_{t-1}). The computation of the dividend to price ratio on the S&P500 index follows Fama and French (1988). The number of observations is 20. *t*-statistics are shown in parentheses.

Intercept	VWD_{t-1}	Div/Price	Adjusted R^2
456.9	−19.3	—	40.9
	(−3.76)		
230.1	−21.8	61.8	41.5
	(−3.90)	(1.09)	

The following figure (authors' Figure 4) buttresses the correlation between IPO activity and closed-end fund discount

Figure 4 The number of IPO's and the discount at the beginning of the year.



This graph shows the number of Initial Public Offerings (IPO's) during the year and the percentage discount on a value-weighted portfolio of closed-end funds at the beginning of the year during 1966 to 1985. The line graph represents the value-weighted discount at the beginning of the year × 50. The bar graph represents the number of IPO's during the year (Source for IPO data: Ibbotson, Sindelar and Ritter (1988)).

Conclusion

Closed-end fund discounts are a measure of the sentiment of individual investors

This investor sentiment is sufficiently widespread to affect the prices of small-cap stocks in the same way it influences the prices of closed-end funds.

Changing investor sentiment makes funds riskier than the portfolios they hold, causing average underpricing of funds relative to fundamentals

Because the same investor sentiment affects small caps and makes them riskier, small caps must also be underpriced relative to their fundamentals—that is, offer a higher expected return as well

This phenomenon is known as the small-firm effect and is discussed in the chapter "Predictability of Stock Returns"

In other words, the authors suggest that the small-firm effect is clientele-related

Note that the finding that the small-firm effect has diminished over the last couple of years fits into the picture of increased institutional ownership in small caps.

[Blank page]

Empirical Studies

11. Stock Market Overreaction (I)

Reference:

De Bondt, Werner F.M., and Richard H. Thaler (1986) "Does the Stock Market Overreact?" *Journal of Finance* 49: 793-807.

> *"… prices have been based too much on current earning power and too little on long-term dividend paying power"*
>
> John Burr Williams
>
> *The Theory of Investment Value,* Amsterdam: North-Holland, 1956 (reprint of the 1938 edition), p. 19.

> *"… the interval required for a substantial underevaluation to correct itself averages approximately 1 1/2 to 2 1/2 years"*
>
> Benjamin Graham
>
> *The Intelligent Investor: A Book of Practical Counsel,* 3rd. ed. New York: Harper & Brother, 1959, p. 37.

Note: All figures and tables in this chapter are from Bondt and Thaler (1986)

The Bond and Thaler (1986) study tries to establish evidence for overreaction in the stock market

Companies with very low P/E (price/earnings) ratios tend to be temporarily "undervalued" because investors have become excessively pessimistic after a series of bad earnings reports or other bad news

Remember that undervaluation means that the company's market value falls short of the intrinsic value

As the company's market value eventually reverts to the intrinsic value (mean reversion; regression to the mean), the company's stock outperforms the market

Mean reversion implies predictability of returns

Remember that such abnormal returns violate the weak form of market efficiency.

To gauge overreaction, the authors use three alternative concepts of return residual, yet report empirical results only for one of them

Sharpe-Lintner CAPM residuals

In the CAPM, the following risk-return relationship holds for any portfolio (stock) j:

$$r_{j,t} - r_{f,t} = \beta_j \cdot (r_{m,t} - r_{f,t}) + \varepsilon_{j,t} \ , \ E[\varepsilon_j] = 0$$

where r_f and r_m are the risk-free rate of the return and the "market" rate of return, respectively

After estimating the parameter β_j econometrically, we can write for the return residual:

$$\hat{\varepsilon}_{j,t} = (r_{j,t} - r_{f,t}) - \hat{\beta}_j \cdot (r_{m,t} - r_{f,t})$$

The residual $\hat{\varepsilon}_j$ is an abnormal return

The CAPM (as well as the ordinary least squares method) implies that, on average, $\hat{\varepsilon}_j$ is zero.

The residual is a beta-adjusted return; it measures the stock's return over and above what one would predict based on market movements in that period, given the stock's sensitivity to the market portfolio.

Market-model residuals

The market model reads

$$r_{j,t} = \alpha_j + \beta_j \cdot r_{m,t} + \varepsilon_{j,t}$$

After estimating the parameters α_j and β_j econometrically, we can write:

$$\hat{\varepsilon}_{j,t} = r_{j,t} - (\hat{\alpha}_j + \hat{\beta}_j \cdot r_{m,t})$$

Again, the residual is an abnormal return; like the CAPM residual return, it is a measure for the stock's return over and above what one would predict based on market movements in that period, given the stock's sensitivity to the market portfolio

The difference to the CAPM is twofold:

The parameter α is set to zero in the CAPM

The CAPM uses excess returns ($r_{j,t} - r_{f,t}$; $r_{m,t} - r_{f,t}$), rather than <u>raw returns</u> ($r_{j,t}$; $r_{m,t}$).

The motivation for the market model is that the Sharpe-Lintner model is <u>highly structured</u>

Highly structured models are most susceptible to the joint hypothesis problem.

Market-adjusted excess returns (the only of the three concepts for which empirical results are reported)

$$\varepsilon_t = r_{j,t} - r_{m,t}$$

The residual is no abnormal return; rather, it is an excess return

The residual measures the stock's return over and above the market.

Note that this approach implicitly assumes that all stocks are equally sensitive to market movements, i.e., that all stocks have a market beta equal to one.

The authors use monthly return data for New York Stock Exchange (NYSE) common stocks for the period January 1926 through December 1982, as compiled by the Center for Research in Security Prices (CRSP) of the University of Chicago

The authors choose 36-month windows (test periods), starting in January 1930, January 1933, …, January 1975

Stocks that qualify must have at least 48 consecutive months of return data prior to the test period available

There is a total of 16 non-overlapping test periods, which are preceded by 36-month portfolio formation periods

For each portfolio formation period, the authors calculate the cumulative market-adjusted excess return

$$\sum_{t=1}^{36} \varepsilon_t = \sum_{t=1}^{36} (r_{j,t} - r_{m,t})$$

The authors rank the stocks by their performance during the portfolio formation period

The top 35 (50; top decile) stocks are assigned to the winner portfolio (W)

The bottom 35 (50; bottom decile) are assigned to the loser portfolio (L).

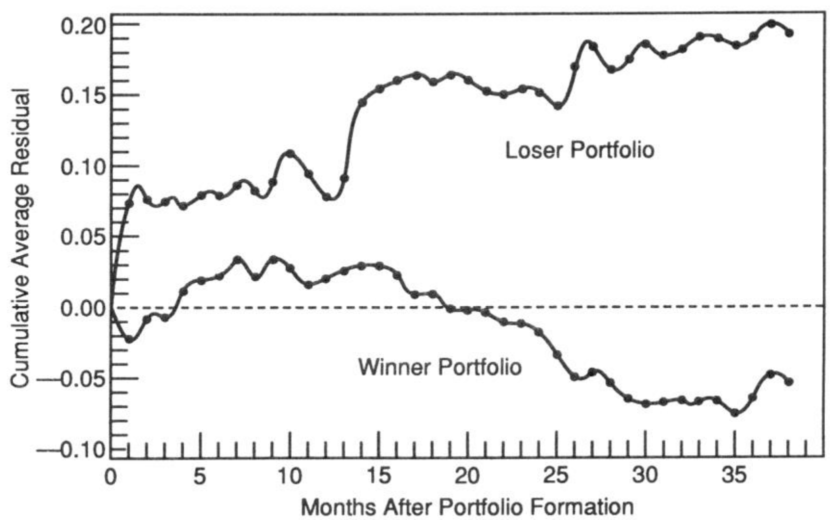
For each of the 36 months of the test period, the authors calculate the market-adjusted excess return of each stock in the winner and loser portfolios, respectively

For each portfolio, the authors average (seemingly without weighting) the market-adjusted excess returns of the individual stocks

The number of stocks in the portfolios W or L might decrease over time as stocks are taken off the ticker.

Cumulative average excess market-adjusted returns

Figure 1 Cumulative Average Residuals for Winner and Loser Portfolios of 35 Stocks (1–36 months into the test period)



36 months after portfolio formation, loser portfolios of 35 stocks have outperformed the market by, on average, 16.9 percentage points

Winner portfolios, on the other hand, have earned about 5 percentage points less than the market

The difference in cumulative abnormal returns between the two portfolios within the 36-month period equals 24.6 percentage points and is statistically significant (*t*-statistic: 2.20)

Also, the overreaction effect is asymmetric in that it is larger for losers than for winners.

Moreover, there is a pronounced "turn-of-the-year effect" ("January effect") in that most of the excess returns are realized in January

In months $t = 1$, $t = 13$, and $t = 25$, the loser portfolio earns excess returns of, respectively, 8.1 percentage points (*t*-statistic: 3.21), 5.6 percentage points (3.07), and 4.0 percentage points (2.76)

The January effect is due to tax-induced selling

If a stock has suffered a severe loss over the calendar year, personal investors might realize the tax loss late in the year before repurchasing the stock after a period of at least 30 days.

Consistent with Graham's claim, the correction mostly occurs during the second and third years of the test period.

The following table shows results for …

… alternative sizes of winner/loser portfolios (50 stocks; deciles)

… alternative lengths of the portfolio formation period (1,2,5 years)

… alternative lengths of the test period (1-60 months)

Table 1 Differences in Cumulative Average (Market-Adjusted) Residual Returns Between the Winner and Loser Portfolios at the End of the Formation Period, and 1, 12, 13, 18, 24, 25, 36, and 60 Months into the Test Period

Portfolio Selection Procedures: Length of the Formation Period and No. of Independent Replications	Average No. of Stocks	CAR at the End of the Formation Period		Difference in CAR (*t*-Statistics)							
				Months After Portfolio Formation							
		Winner Portfolio	Loser Portfolio	1	12	13	18	24	25	36	60
10 five-year periods	50	1.463	−1.194	0.070 (3.13)	0.156 (2.04)	0.248 (3.14)	0.256 (3.17)	0.196 (2.15)	0.228 (2.40)	0.230 (2.07)	0.319 (3.28)
16 three-year periods	35	1.375	−1.064	0.105 (3.29)	0.054 (0.77)	0.103 (1.18)	0.167 (1.51)	0.181 (1.71)	0.234 (2.19)	0.246 (2.20)	NA*
24 two-year periods[a]	35	1.130	−0.857	0.062 (2.91)	−0.006 (−0.16)	0.074 (1.53)	0.136 (2.02)	0.101 (1.41)	NA	NA	NA
25 two-year periods[b]	35	1.119	−0.866	0.089 (3.98)	0.011 (0.19)	0.092 (1.48)	0.107 (1.47)	0.115 (1.55)	NA	NA	NA
24 two-year periods[a] (deciles)	82	0.875	−0.711	0.051 (3.13)	0.006 (0.19)	0.066 (1.71)	0.105 (1.99)	0.083 (1.49)	NA	NA	NA
25 two-year periods[b] (deciles)	82	0.868	−0.714	0.068 (3.86)	0.008 (0.19)	0.071 (1.46)	0.078 (1.41)	0.072 (1.29)	NA	NA	NA
49 one-year periods	35	0.774	−0.585	0.042 (2.45)	−0.076 (−2.32)	−0.006 (−0.15)	0.007 (0.14)	−0.005 (−0.09)	NA	NA	NA

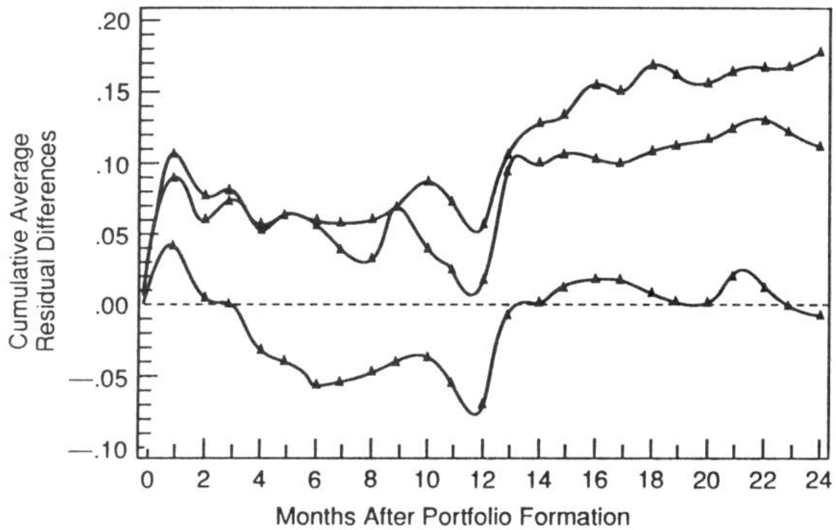
a. The formation month for these portfolios is the month of December in all uneven years between 1933 and 1979.

b. The formation month for these portfolios is the month of December in all even years between 1932 and 1980.

c. NA, not applicable.

Note that for a formation period of one year (last row of table), no reversal is observed, which is confirmed in the following figure

Figure 2 Differences in Cumulative Average Residual Between Winner and Loser Portfolios of 35 Stocks (formed over the previous one, two, or three years; 1–24 months into the test period)



The upper (middle) line indicates a portfolio formation period of 36 (24) months

Note that for a formation period of one year (last row of table), there is <u>momentum</u>

The lower portfolio remains a loser over the first year of the test period

Generally, momentum might indicate <u>initial underreaction</u> or it might indicate that <u>overreaction takes time to run its course</u>

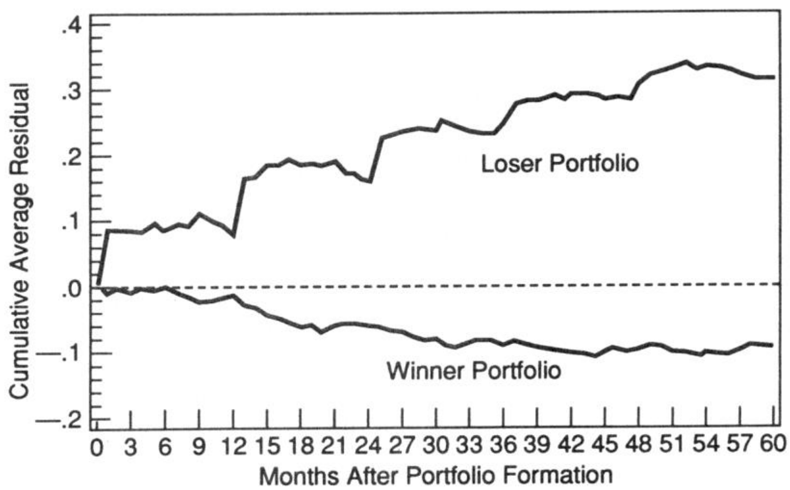
The evidence provided in the Bondt and Thaler study does not allow one to distinguish between the two possible causes.

Are the results due to risk?

- Remember that the concept of the market-adjusted excess return implies a beta of unity
- Is the difference in performance between the two portfolios an artifact of the winner portfolio having higher systematic risk (a higher market beta) than the loser portfolio?
  - The authors estimate the CAPM betas of the portfolios <u>for the portfolio formation period</u> and find that for the portfolios displayed in Figure 1, the beta of the winner portfolio equals 1.369, while the beta of the loser portfolios amounts to only 1.026 (*t*-statistic: 3.09)
    - (The *t*-test might be incorrect because the estimated betas are stochastic)
    - The authors conclude that the market-adjusted approach <u>underestimates</u> the magnitude of the mean reversion (and thus the initial overreaction)
      - The authors fail to recognize that what matters is the beta during the test period, rather than the beta of the portfolio formation period.

Finally, the authors look into a trading rule where the investor forms a loser portfolio (at the end of) every December between 1932 and 1977 (based on a the performance in the prior five years) and holds on to each of these portfolios for five years

Figure 3 Cumulative Average Residuals for Winner and Loser Portfolios of 35 Stocks (1–60 months into the test period)



Conclusion

Investors tend to overreact to strings of bad news, which causes undervaluation of losers and subsequent mean reversion

Portfolios that have been beaten down over the last two or three calendar years create abnormal returns over the next three years.

Before mean reversion sets in, there is momentum in stocks

Portfolios that have been losers during the last calendar year stay losers for another calendar year.

Momentum as well as mean reversion makes returns predictable.

Criticism

K.C. Chan (1988, "On the Contrarian Investment Strategy," *Journal of Business* 16, 147-163) and R. Ball and S.P. Kothari (1991, "Security Returns around Earnings Announcements," *The Accounting Review* 66, 718-738) argue that the return reversals evidenced by De Bondt and Thaler are due primarily to systematic changes in equilibrium-required returns

Ball and Kothari report that the betas of extreme losers exceed the betas of extreme winners by 0.76 following the portfolio formation period (i.e., during the test period)

- Given historical risk premiums $(r_{m,t} - r_{f,t})$, this difference can explain a great deal of the differences in realized returns between the winner and loser portfolios
- Remember that De Bondt and Thaler look at the betas of the portfolio formation period, not at the betas of the test period.

P. Zarowin (1990, "Size, Seasonality, and Stock Market Overreaction," *Journal of Financial and Quantitative Analysis* 25, 113-125) argues that the return reversals found by De Bondt and Thaler might be due to the small firm effect

Small firms are over-represented in loser portfolios.

Empirical Studies

12. Stock Market Overreaction (II)

Reference:

Chopra, Navin, Josef Lakonishok and Jay R. Ritter (1992) "Measuring Abnormal Performance: Do Stocks Overreact?" *Journal of Financial Economics* 31, 235-268.

Note: All figures and tables in this chapter are Chopra, Lakonishok and Ritter (1992)

The Chopra, Lakonishok and Ritter (1992) study is a follow-up to Bond and Thaler (1986)

Chopra, Lakonishok and Ritter try to correct (what they think are) deficiencies of the De Bondt and Thaler study

K.C. Chan (1988, "On the Contrarian Investment Strategy," *Journal of Business* 16, 147-163) and R. Ball and S.P. Kothari (1991, Security Returns around Earnings Announcements, *The Accounting Review* 66, 718-738) argue that the return reversals evidenced by De Bondt and Thaler are due primarily to systematic changes in equilibrium-required returns

Ball and Kothari report that the betas of extreme losers exceed the betas of extreme winners by 0.76 following the portfolio formation period

Given historical risk premiums $(r_{m,t} - r_{f,t})$, this difference can explain a great deal of the differences in realized returns between the winner and loser portfolios

Remember that De Bondt and Thaler look at the betas of the portfolio formation period, not at the betas of the test period.

P. Zarowin (1990, Size, Seasonality, and Stock Market Overreaction, *Journal of Financial and Quantitative Analysis* 25, 113-125) argues that the return reversals found by De Bondt and Thaler might be due to the small firm effect

Small firms are over-represented in loser portfolios.

Empirical methodology

The authors use monthly data for New York Stock Exchange (NYSE) common stocks from 1926 to 1986, as compiled by the Center for Research in Security Prices (CRSP) of the University of Chicago

Only stocks that have been listed continuously for a period of five calendar years are included; this five-year period is called the ranking period (year −4 to year 0; the ranking period compares to the portfolio formation period in De Bondt and Thaler (1986))

The stocks are ranked into twenty portfolios, based on their buy-and-hold raw returns during the five-year ranking period

The first ranking period ends in December 1930, and the last one ends in December 1981

There is a total of 52 (overlapping) ranking periods.

The post-ranking periods (year +1 to year +4; the post-ranking period compares to the test period in De Bondt and Thaler) are overlapping five-year intervals starting with 1931-35 and ending with 1982-86

To avoid a survivorship bias, the authors do not require that the stock remains listed for the whole post-ranking period

In the sample in question, approximately 22 percent of the companies in the extreme loser portfolio are delisted during the post-ranking period, while the corresponding number for the extreme winner portfolio is only 8 percent

Delistings are mainly due to bankruptcies and takeovers.

For each of the twenty portfolios, the authors obtain a time series of 52 portfolio returns for each of the ten event years (five-year ranking period plus five-year post-ranking period)

For each of the twenty portfolios, $p = 1,...,20$, and each of the ten event years, $\tau = 1,...,10$, the authors estimate the following model based on 52 annual observations:

$$r_{p,\tau} - r_f = \alpha_{p,\tau} + \beta_{p,\tau} \cdot (r_m - r_f) + \varepsilon_{p,\tau} \text{ , } E[\varepsilon_{p,\tau}] = 0$$

where r_f and r_m are taken from the respective calendar years

The model differs from the Sharpe-Lintner CAPM in that it contains a parameter α, called Jensen's alpha

If a portfolio p outperforms (underperforms) the market, its α is greater (smaller) than zero (and statistically significant)

Jensen's alpha is a direct measure of an abnormal return.

The market portfolio, m, is alternatively the equally-weighted (EW) market return on NYSE stocks or the corresponding value-weighted (VW) market return

The risk-free rate, r_f, is the annual return on Treasury bills (T-bills) (possibly 30-day maturities, rolled over).

Table 1 Average annual post-ranking-period percentage returns, alphas, and betas for twenty portfolios formed on the basis of either ranking-period returns or ranking-period betas. Average monthly post-ranking-period percentage returns, alphas, and betas are also reported for portfolios formed on the basis of five-year ranking-period returns.

Alphas and betas are estimated from time-series regressions with 52 observations, for ranking periods ending in 1930–81, for each of the five post-ranking years. The alphas and betas reported are the averages of these five post-ranking-period numbers. In columns (1)–(5) and (9)–(11), portfolio 1 is comprised of stocks with the lowest ranking-period returns, and portfolio 20 is comprised of the stocks with the highest ranking-period returns. In columns (6)–(8), portfolio 1 is comprised of stocks having the highest ranking-period betas, and portfolio 20 is comprised of stocks having the lowest ranking-period betas. EW and VW are, respectively, equally-weighted and value-weighted market indices of NYSE stocks. Columns (1)–(8) are based upon annual returns, whereas columns (9)–(11) are based upon monthly returns.

	Portfolios Formed on the Basis of Ranking-Period Returns					Portfolios Formed on the Basis of Ranking-Period Betas			Portfolios Formed on the Basis of Ranking-Period Returns		
		Computed Using EW Index		Computed Using VW Index			Computed Using EW Index			Computed Using EW Index	
Portfolio	Average Annual Return (%) (1)	Alpha (2)	Beta (3)	Alpha (4)	Beta (5)	Average Annual Return (%) (6)	Alpha (7)	Beta (8)	Average Monthly Return (%) (9)	Alpha (10)	Beta (11)
1	27.3	−0.2	1.65	2.7	1.95	21.0	−3.0	1.42	2.36	0.26	1.52
2	23.0	0.5	1.31	2.5	1.62	19.4	−3.5	1.34	1.90	0.07	1.30
3	21.0	0.1	1.20	1.9	1.51	20.3	−1.2	1.25	1.80	0.05	1.23
4	21.2	0.9	1.16	2.9	1.45	20.4	−0.8	1.22	1.73	0.09	1.14
5	20.5	1.2	1.09	2.8	1.39	21.0	−0.6	1.24	1.65	0.09	1.07
6	19.9	0.7	1.08	2.2	1.40	20.2	0.0	1.15	1.59	0.06	1.05
7	19.4	0.0	1.09	1.6	1.40	20.2	0.1	1.14	1.52	0.01	1.03
8	18.5	1.5	0.94	2.9	1.24	19.8	0.1	1.12	1.48	0.06	0.95
9	17.6	0.2	0.95	1.7	1.26	18.5	−0.2	1.04	1.41	−0.03	0.98
10	17.8	0.7	0.94	2.1	1.24	18.3	−0.5	1.06	1.43	0.03	0.94
11	16.9	0.2	0.91	1.4	1.22	19.3	0.5	1.05	1.35	−0.04	0.93
12	16.6	0.1	0.89	1.2	1.22	17.3	0.0	0.95	1.34	−0.03	0.92
13	16.7	0.2	0.90	1.4	1.22	17.2	0.4	0.91	1.33	−0.00	0.88
14	16.1	−0.2	0.88	0.8	1.21	17.2	0.8	0.89	1.29	−0.06	0.90
15	15.5	−0.2	0.84	0.9	1.16	16.2	0.9	0.82	1.25	−0.07	0.87
16	15.3	−0.6	0.85	0.3	1.18	15.4	0.6	0.78	1.20	−0.07	0.83
17	14.6	0.1	0.76	1.0	1.08	15.2	1.4	0.72	1.16	−0.05	0.78
18	14.5	−1.3	0.85	−0.5	1.18	14.2	1.7	0.62	1.10	−0.12	0.79
19	14.3	−1.3	0.84	−0.7	1.19	14.4	1.1	0.67	1.11	−0.12	0.79
20	13.3	−2.7	0.86	−2.0	1.21	13.7	2.1	0.56	1.01	−0.24	0.81
Mean	18.0	0.0	1.00	1.35	1.32	18.0	0.0	1.00	1.45	−0.06	0.98
$r_1 - r_{20}$	14.0	2.5	0.79	4.7	0.74	7.3	−5.1	0.86	1.35	0.50	0.71

Table1, columns (1)-(5) presents the empirical results of the estimated model

- The annual raw returns, on the basis of which the portfolios were formed in the five-year ranking period, are shown in column (1), averaged over the 52 moving windows
- Estimates of the alphas and betas for the equally-weighted portfolio (EW) are displayed in columns (2)-(3), while estimates for the alphas and betas of the value-weighted (VW) portfolio are shown in columns (4)-(5)
    - For each of the twenty portfolios, the alphas and betas were averaged over the five-year post-ranking period.
- The difference in raw returns—which are greater than zero for all portfolios—between the extreme winner and loser portfolios is shown as $r_1 - r_{20} = 14.0$ percentage points per annum
    - Added up over the five-year post-ranking period even before compounding, the difference in returns amounts to 70 percentage points

The difference in betas between the extreme winner (w) and loser (l) portfolios amounts to 0.79

Given an equity risk premium, $r_f - r_m$, in the range of $14 - 15$ percentage points using the equally-weighted portfolio (for which the mean alpha is zero), the CAPM predicts a difference in returns of approximately 11 percent:

$$r_w - r_l = (r_w - r_f) - (r_l - r_f)$$

$$= \hat{\beta}_w \cdot (r_{m,t} - r_{f,t}) + \hat{\varepsilon} - \hat{\beta}_l \cdot (r_{m,t} - r_{f,t}) - \hat{\varepsilon}$$

$$= (\hat{\beta}_w - \hat{\beta}_l) \cdot (r_{m,t} - r_{f,t}) = 0.145 \times 0.79 \approx 11.5 \text{ percent}$$

Consequently, only 2.5 percentage points of the 14 percentage point difference in the performance of the extreme winner and loser portfolios are due to abnormal returns

As shown in column (2), the difference in Jensen's alphas between the extreme winner and loser portfolios is indeed 2.5 percentage points.

Numerous studies have found that the securities market line is flatter than implied by the slope $r_f - r_m$, which is the unit price of systematic risk in the CAPM (the equity (market) risk premium)

If the empirical equity risk premium is indeed lower than implied by the CAPM, the measured difference in the alphas between the loser and the winner portfolios underestimates the actual difference in abnormal returns

Instead of deriving the market compensation per unit of systematic risk from the CAPM, the unit price of systematic risk can be estimated empirically by regressing the averaged twenty post-ranking period raw returns, $\bar{r}_p$, on the averaged betas, $\bar{\hat{\beta}}_p$ shown in Table 1:

$$\bar{r}_p = a + b \cdot \bar{\hat{\beta}}_p + \varepsilon_p \ , \ E[\varepsilon_p] = 0$$

The authors employ this methodology, using portfolios that were formed based on the ranking-period betas (rather than the ranking-period raw returns, which were used for columns (1)-(5) in Table 1

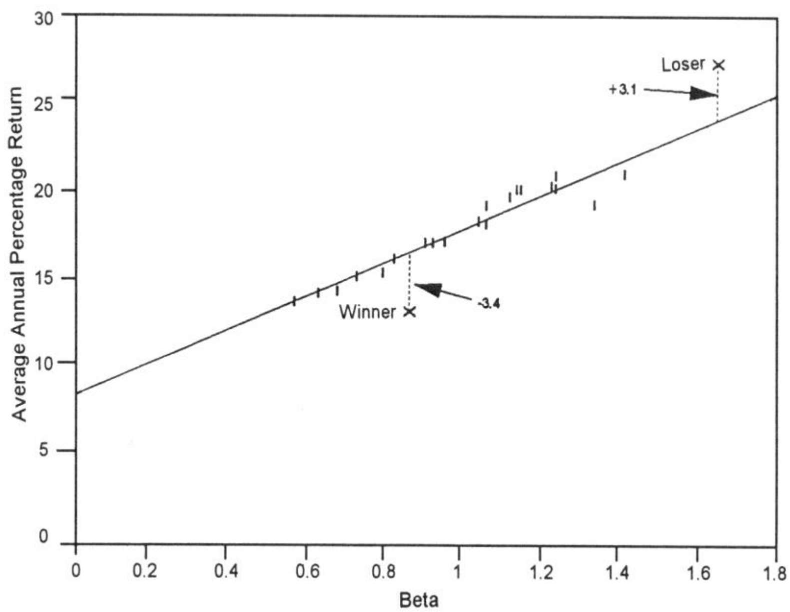
First, alpha and beta are estimated for each stock: The stock is then ranked on the basis of the estimated beta into one of twenty portfolios

The averaged estimated post-ranking period alphas and betas of these twenty portfolios, along with the raw returns, are shown in columns (6)-(8) in Table 1.

Second, the market compensation per unit of systematic risk is estimated using the data on averaged raw returns and averaged betas displayed in columns (6) and (8) of Table 1

The regression results are displayed in Figure 1a.

Figure 1a Plot of the empirical security market line (SML) calculated using annual data from the realized post-ranking-period returns and betas for twenty portfolios formed on the basis of ranking-period betas, and the realized post-ranking-period return on extreme winner and loser portfolios.



The empirical SML is estimated from the twenty portfolio returns and betas reported in columns (6) and (8) of table 1. The empirical SML has an intercept of 8.5% and a slope of 9.5%. Alphas are calculated as deviations from the empirical SML.

The regression equation has an intercept of 8.5 percent and a slope (unit price of systematic risk) of 9.5 percent

- The 8.5 percent intercept is markedly higher than the average risk-free rate of return in the sample period of about 3.5 percent
- The slope coefficient of 9.5 percent is considerable lower than the 14-15 percent equity (market) risk premium
  - Differences in betas do not generate differences in returns during the sample period as great as assumed by the Sharpe-Lintner CAPM
  - This finding is consistent with other empirical studies on the unit price for systematic risk.

According to Figure 1a, the difference in abnormal return between the extreme winner and loser portfolios (marked in the figure) amounts to 6.5 percentage points

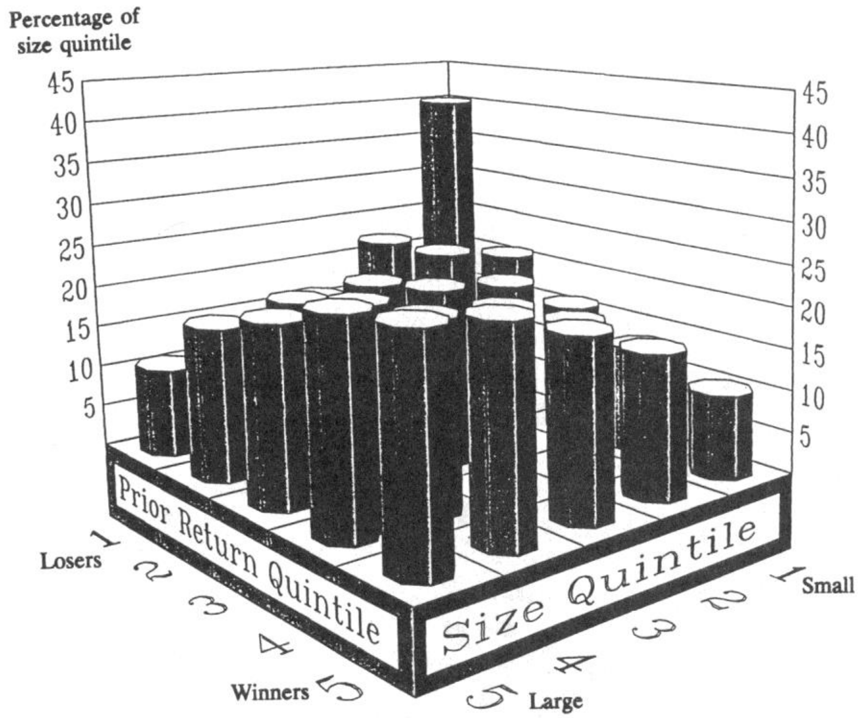
> The abnormal return of 6.5 percentage points based on the empirical securities market line exceeds considerably the 2.5 percentage points derived theoretically from the Sharpe-Lintner CAPM.

Is the measured abnormal return difference between the extreme loser and winner portfolios a truly risk-adjusted return, or is an artifact of a neglected risk factor?

- Zarowin (1990) has shown that losers have lower market capitalizations than winners
  - If firm size is a risk factor—as assumed in multi-factor models—then overrepresentation of small firms in loser portfolios may explain why loser portfolios outperform the market.

Figure 2 plots the joint distribution of firms categorized by size and prior returns

Figure 2 **The joint distribution of firms categorized by size and prior returns.**

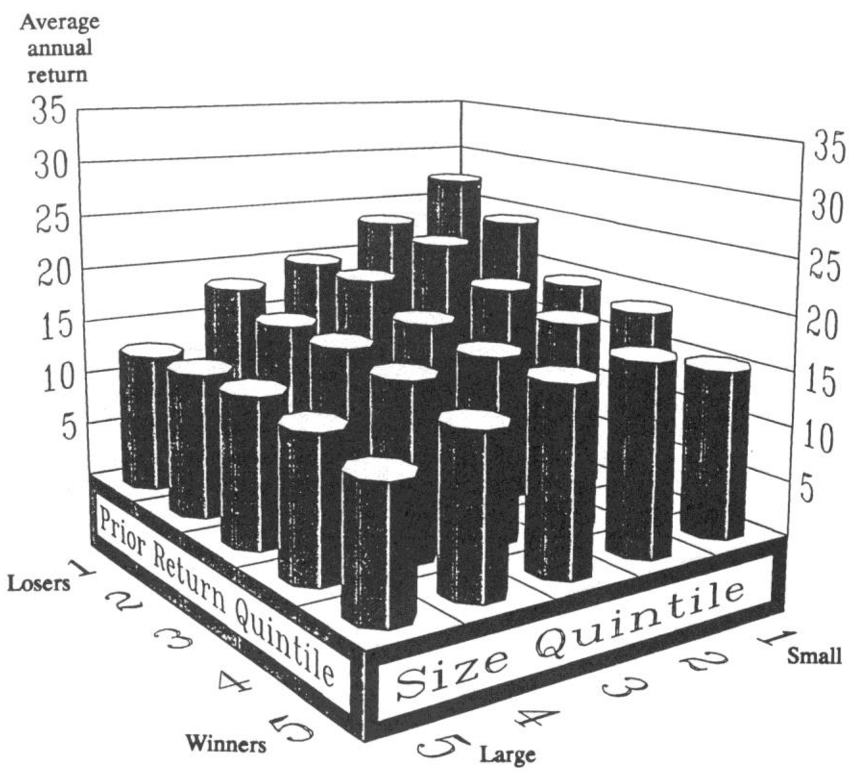


For each size quintile, the percentage of firms falling in each prior return quintile is plotted. Quintile portfolios are plotted rather than the twenty portfolios used in the empirical work because 400 portfolios (20 × 20) produces too cluttered a figure compared with the 25 portfolios plotted.

Figure 2 shows that in the smallest size quintile, 40 percent of the firms are in the extreme loser quintile, while only 10 percent are in the extreme winner quintile.

Figure 3 plots the joint distribution of average annual raw returns in the post-ranking period, categorized by size and prior returns

Figure 3 The joint distribution of average annual returns in the post-ranking period categorized by size and prior returns.



The average annual return on the smallest quintile of losers is 27.37%, while the average annual return on the largest quintile of winners is 11.59%.

On average, size constant, the extreme loser quintile has a 5.4 percentage point higher average annual return than the extreme winner quintile

On average, prior returns constant, the smallest size quintile has an 8.2 percentage point higher average annual return than the largest size quintile.

Table 2 reports differences in size-adjusted returns between extreme winner and loser portfolios

Table 2 Average annual post-ranking-period percentage returns for twenty portfolios of firms ranked by their five-year ranking-period returns, size-control portfolios with and without losers and winners purged, and the associated size-adjusted returns.

The twenty size-control portfolios are constructed to have approximately the same market values as the twenty ranked portfolios. Excess returns are computed two different ways: (i) size-adjusted returns using all firms (unpurged) and (ii) size-adjusted returns after the portfolios have been purged of all firms in the top five and the bottom five portfolios of prior returns (purged).

	Average Annual Return (%) in Years +1 to +5				Size-Adj. Returns (%) $e = r_p - r_s$	
		Control Firms				
Portfolio	Ranked Firms (r_p) (1)	Unpurged (r_s) (2)	Purged (r_s) (3)	Difference (2)–(3) (4)	Unpurged (1)–(2) (5)	Purged (1)–(3) (6)
1	27.3	23.4	20.4	3.0	3.9	6.9
2	23.0	21.3	19.3	2.0	1.7	3.7
3	21.0	20.6	19.0	1.6	0.4	2.0
4	21.2	20.0	18.8	1.2	1.2	2.4
5	20.5	19.4	18.0	1.4	1.1	2.5
6	19.9	18.8	18.0	0.8	1.1	1.9
7	19.4	18.9	18.1	0.8	0.5	1.3
8	18.5	18.1	17.6	0.5	0.4	0.9
9	17.6	17.9	17.4	0.5	−0.3	0.2
10	17.8	17.5	17.2	0.3	0.3	0.6
11	16.9	17.3	16.9	0.4	−0.4	0.0
12	16.6	17.0	16.7	0.3	−0.4	−0.1
13	16.7	16.9	16.8	0.1	−0.2	−0.1
14	16.1	16.6	16.3	0.3	−0.5	−0.2
15	15.5	16.6	16.4	0.2	−1.1	−0.9
16	15.3	16.6	16.4	0.2	−1.3	−1.1
17	14.6	16.2	16.1	0.1	−1.6	−1.5
18	14.5	16.0	16.1	−0.1	−1.5	−1.6
19	14.3	16.0	15.9	0.1	−1.7	−1.6
20	13.3	16.0	16.1	−0.1	−2.7	−2.8
Mean	18.0	18.0	17.4	0.06	0.0	0.6
$r_1 - r_{20}$	14.0	7.4	4.3	3.1	6.6	9.7

Column (1) of Table 2 is identical to column (1) of Table 1; it shows the ranking-period raw returns of twenty portfolios ranked by raw returns (rather than betas, as done in column (6) of Table 1)

- Remember that the difference in raw returns between the extreme loser and winner portfolios equals 14.0 percentage points ($r_1 - r_{20}$; Table 2, column (1), last row).

Column (2) shows the returns on control portfolios formed by matching size (size-control porfolios)

- To construct the size-control portfolios, the authors rank the population of firms at the end of the portfolio formation period by market capitalization and assign them to one of twenty size portfolios
- The authors then calculate for each of the twenty portfolios ranked by raw returns which fraction of firms falls into each size portfolio
    - These fractions serve as weights when calculating the raw return of the size-control portfolio.
- The return of the size-control portfolio shown in column (2) is a weighted average of the returns of all twenty portfolios, ranked by size.

In column (3) the authors report the average annual return on “purged” size-control portfolios

- These portfolios have been formed in a manner identical to that employed in column (2), with the exception that the population of firms from which the size portfolios are drawn has been purged of firms in prior return portfolios 1-5 (losers) and 16-20 (winners).

In column (5), the authors report excess returns computed by subtracting the unpurged size-control returns (column (2)) from the raw returns (column (1))

- There is a nearly monotonic decrease in the excess returns as one goes from portfolio 1 (extreme losers) to portfolio 20 (extreme winners)
- The difference in excess returns between the extreme loser and winner portfolios averages 6.6 percentage points per annum during the five post-ranking years
  - When size is controlled for without account for the correlation of size and prior returns, mean reversion might be underestimated.

In column (6), the authors report excess returns computed by subtracting the purged size-control returns (column (3)) from the raw returns (column (1))

- The results accentuate the findings for the unpurged size-control portfolios
- The difference in excess returns between the extreme loser and winner portfolios averages 9.7 percentage points per annum during the five post-ranking years.

Seasonal return patterns

Table 3 reports average post-ranking period raw returns and (unpurged) size-adjusted returns …

… on an annual basis

… for the month of January (not annualized)

… for the period of February-December (11-month compounded returns).

Table 3 Seasonal patterns in raw returns and size-adjusted returns for ranking periods of five years and one year.[a]

	Five-Year Ranking Periods						One-Year Ranking Periods					
	Av. Raw Returns (%)			Avg. Size-Adj. Returns[b] (%)			Avg. Raw Returns (%)			Avg. Size-Adj. Returns[b] (%)		
Portfolio	Annual (1)	Jan. (2)	Feb.–Dec. (3)	Annual (4)	Jan. (5)	Feb.–Dec. (6)	Annual (7)	Jan. (8)	Feb.–Dec. (9)	Annual (10)	Jan. (11)	Feb.–Dec. (12)
1	27.3	13.1	12.9	6.9	7.2	−0.8	23.5	11.2	11.3	2.6	4.7	−2.2
2	23.0	8.8	13.3	3.7	3.5	0.1	20.5	7.4	12.0	1.3	1.9	−1.0
3	21.0	7.4	12.9	2.0	2.5	−0.4	19.8	6.6	12.4	0.9	1.4	−0.5
4	21.2	6.6	13.8	2.4	1.7	0.7	18.4	5.9	11.9	0.1	0.8	−0.6
5	20.5	5.7	14.0	2.5	1.2	1.2	18.7	5.5	12.6	1.0	0.8	0.4
6	19.9	5.6	13.5	1.9	1.2	0.5	18.1	5.3	12.1	0.3	0.7	−0.4
7	19.4	5.1	13.6	1.3	0.9	0.5	17.5	5.0	12.0	−0.3	0.4	−0.5
8	18.5	4.7	13.2	0.9	0.5	0.5	18.2	4.6	12.8	0.7	0.2	0.3
9	17.6	4.5	12.4	0.2	0.4	−0.2	16.8	4.4	11.8	−0.4	0.1	−0.5
10	17.8	4.2	13.0	0.6	0.2	0.5	17.5	4.0	12.8	0.2	−0.2	0.4
11	16.9	4.1	12.3	0.0	0.1	0.1	16.9	4.3	12.3	−0.6	0.0	−0.2
12	16.6	3.9	12.2	−0.1	0.0	−0.1	17.6	4.0	12.8	0.3	−0.1	0.3
13	16.7	3.6	12.6	−0.1	−0.1	0.2	17.0	3.9	12.5	0.0	−0.3	0.3
14	16.1	3.3	12.3	−0.2	−0.3	0.3	16.7	3.8	12.3	−0.5	−0.4	0.0
15	15.5	3.3	11.8	−0.9	−0.3	−0.4	16.8	3.6	12.7	0.0	−0.4	0.5
16	15.3	3.2	11.4	−1.1	−0.4	−0.8	16.9	3.7	12.6	−0.1	−0.3	0.3
17	14.6	3.1	11.0	−1.5	−0.5	−0.9	16.9	3.5	12.7	0.1	−0.6	0.7
18	14.5	3.0	10.8	−1.6	−0.4	−1.2	16.3	3.6	12.2	−1.2	−0.7	−0.3
19	14.3	2.7	10.9	−1.6	−0.7	−1.0	17.5	3.7	13.1	−0.4	−0.8	0.5
20	13.3	2.6	10.0	−2.8	−0.7	−2.1	17.7	4.3	12.8	−0.9	−0.6	−0.1
$r_1 - r_{20}$	14.0	10.5	2.9	9.7	7.9	1.3	5.8	6.9	−1.5	3.5	5.3	−2.1
p-values[c]	0.001	0.001	0.105	0.005	0.004	0.273	0.001	0.001	0.072	0.004	0.001	0.024
$r_1 - r_{20}$ in year +1[d]	15.3	15.1	−0.7	11.0	12.6	−2.6	−7.2	8.6	−15.1	−8.6	7.2	−15.2
p-values[e]	0.011	0.000	0.425	0.028	0.000	0.176	0.013	0.000	0.000	0.004	0.000	0.000

Tests of the hypothesis that $r_1 - r_{20}$ with five-year ranking periods = $r_1 - r_{20}$ with one-year ranking periods:

	(1)	(2)	(3)	(4)	(5)	(6)
p-values[f]	0.002	0.003	0.007	0.009	0.035	0.037

a. All numbers, except for the row labeled '$r_1 - r_{20}$ in year +1', are the equally-weighted averages of the five post-ranking years, for all 52 post-ranking periods beginning in 1931–1982. The January returns are monthly averages. The February–December returns are averages of eleven-month compounded returns.

b. The size-control portfolios have been purged of extreme winners and losers, using the procedures described in table 2. The purged firms for the one-year ranking periods are those in the bottom 25% and the top 25% of one-year returns.

c. The p-values test the hypothesis that the mean value of $r_1 - r_{20}$ is zero; p-values are computed adjusting for fourth-order autocorrelation as follows, and the standard deviation of the mean value of $r_1 - r_{20}$ is computed as

$$\text{s.d.} = \frac{\sigma}{T}\sqrt{T + 2(T-1)\rho_1 + 2(T-2)\rho_2 + 2(T-3)\rho_3 + 2(T-4)\rho_4},$$

with $T = 52$, where σ is the standard deviation of the portfolio returns and ρ_n is the estimated nth-order simple autocorrelation coefficient. (Four lags are used because of the five-year overlapping post-ranking periods.) The T observations are the time series of five-year average portfolio returns, expressed as annual numbers.

d. The numbers in this row represent the average value of $r_1 - r_{20}$ in the first year of the five post-ranking years.

e. The p-values test the hypothesis that the mean value of $r_1 - r_{20}$ is zero. A time series of 52 nonoverlapping year + 1 observations are used to calculate the standard deviation of the mean, adjusting for first-order autocorrelation. The autocorrelation coefficients are as high as 0.406 for the size-adjusted January returns in column (5). For February–December returns, the autocorrelations are insignificantly different from zero.

f. The p-values are calculated from a time-series of 52 values of $(r_1 - r_{20})_{5,t} - (r_1 - r_{20})_{1,t}$, adjusted for fourth-order autocorrelation, where $(r_1 - r_{20})_{i,t}$ is the average return difference over the five-year post-ranking period starting in year t with ranking period of length i.

In Table 3, columns (1)-(6), the portfolios are formed on the basis of five-year prior returns

- Column (1) is identical to column (1) in Tables 1 and 2
- Column (4), which shows the size-adjusted raw returns, is identical to column (6) in Table 2.

In Table 3, columns (7)-(12), the portfolios are formed on the basis of one-year prior returns (a one-year ranking period)

- Column (1) is identical to column (1) in Tables 1 and 2
- Column (4), which shows the size-adjusted raw returns, is identical to column (6) in Table 2.

Inspection of columns (1)-(6) in Table 3 shows that the overreaction effect is concentrated in January, consistent with the graphical evidence provided in De Bondt and Thaler (1985), Fig. 3

- The differences in average annual and January returns between the extreme winner and loser portfolios ($r_1 - r_{20}$) are statistically significant (see the displayed *p*-values) for both raw returns (Table 3, columns (1) and (2)) and risk-adjusted returns (columns (4) and (5))
- The corresponding difference in average February-December returns between the extreme winner and loser portfolios are not statistically different from zero (see Table 3, column (3), for raw returns and column (6) for risk-adjusted returns).

Is the January effect due to tax-loss selling?

- If a stock has fallen sharply below the purchase price, individual investors tend to realize the loss at the end of the calendar year, just to buy back the stock in January after the mandatory 30-day waiting period

- If the January effect is due to tax-loss selling, the effect might be strongest when portfolios are based on a one-year ranking period

  - Columns (7)-(12) in Table 3 show the seasonal breakdown of raw and risk-adjusted returns for one-year ranking periods

    - The return reversal for the one-year post-ranking period is smaller than for portfolios based on five-year ranking periods

      - Using annual size-adjusted returns, the difference in returns between the extreme winner and loser portfolios is 9.7 percent per year using five-year ranking periods, but only 3.5 percent per year using one-year ranking periods (see $r_1 - r_{20}$ in columns (4) and (10))

      - The return reversal for January, as a fraction of the total annualized return reversal, is indeed higher for one-year ranking periods (in fact, this fraction exceeds 100 percent; see $r_1 - r_{20}$ in columns (7)-(8) and (11)-(12)).

    - There is <u>momentum</u> (negative values of $r_1 - r_{20}$; statistically significant) in the first year of the five-year post-ranking period (year $+1$)

      - Focusing on size-adjusted returns, in the first post-ranking year, prior <u>five-year</u> ranking period losers outperform winners by an annual 11 percentage points (column (4)), whereas prior <u>one-year</u> ranking period losers <u>underperform</u> winners by an annual 15.2 percentage points (column (10))

In other words, when winners and loser are chosen on the basis of one-year returns, losers continue to lose and winners continue to win during the next year

> Similar evidence for momentum has been reported in De Bondt and Thaler (1986) as reported in the chapter “Stock Market Overreaction (I).”

On momentum in stock prices see the chapter “Stock Market Underreaction (Drift).”

Multivariate regression analysis

Multivariate regression allows the authors to simultaneously control for the influence of firm size, prior return, and systematic risk (beta)

For each of the 52 ranking periods, …

… the companies are, independently, ranked into 20 size categories by market value at the end of the five-year ranking period …

… and ranked into 20 return categories based on the five-year ranking period (raw) return

… there are 400 cells, which might not all be of the same number of firms n_k , $k = 1,...,400$.

<u>First</u>, for each of the 400 cells (portfolios, $p = 1,...400$), the market beta is estimated from a pooled (across post-ranking years, $\tau = 1,...,5$, and companies, $j = 1,...,n_k$) regression:

$$r_{j,\tau} - r_f = \alpha + \beta \cdot (r_m - r_f) + \varepsilon, \ E[\varepsilon] = 0$$

where r_f and r_m are taken from the respective calendar years.

<u>Second</u>, the portfolio's excess return is regressed on the portfolio's size category, ranking-period raw return category, and beta:

$$r_p - r_f = a_0 + a_1 \cdot SIZE_p + a_2 \cdot RETURN_p$$
$$+ a_3 \cdot Beta_p + e_p \ , \ E[e_p] = 0$$

Table 4 OLS regressions of average percentage excess returns for the first five post-ranking years for portfolios of NYSE firms formed on the basis of size and prior returns.

For each of the 52 ranking periods ending on December 31 of 1930 to 1981, firms are independently ranked on the basis of their December 31 market value and their five-year prior return, and assigned to one of 400 portfolios. Each portfolio beta is the pooled (over firms and post-ranking years) beta for the firms in the cell, calculated using annual returns and equally-weighted market returns. *SIZE* is measured as the portfolio ranking (1 to 20, with 1 being smallest), and *RETURN* is measured as the portfolio ranking (1 to 20, with 1 being the most extreme prior losers). In panels C and D, *DS* is a dummy variable equal to one if a portfolio is among the bottom 40% of *SIZE* vitiles, *DM* is a dummy variable equal to one if a portfolio is among *SIZE* portfolios 9 to 16 (the middle 40%), and *DL* is a dummy variable equal to one if a portfolio is among the largest 20% of *SIZE* portfolios. *T*-statistics are in parentheses. These are computed using a Fama–MacBeth (1973) procedure adjusted for fourth-order autocorrelation as follows: the *t*-statistic for coefficient a_i is computed as a_i/s.e., where

$$\text{s.e.} = \frac{\sigma}{T}\sqrt{T + 2(T-1)\rho_1 + 2(T-2)\rho_2 + 2(T-3)\rho_3 + 2(T-4)\rho_4},$$

with $T = 52$, where σ is the time-series standard deviation of the coefficient estimates and ρ_t is the estimated *n*th-order simple autocorrelation coefficient. (Four lags are used because of the five-year overlapping post-ranking periods.) The *T* observations are the time series of cross-sectional regression coefficients. The first-order autocorrelations in panel A vary from 0.142 for the intercept to 0.649 for the coefficient on *RETURN*. The R^2 values are based upon the pooled regressions, and do not reflect the year-to-year variability in the regressions.

$$r_p - r_f = a_0 + a_1 SIZE_p + a_2 RETURN_p + a_3 Beta_p + e_p$$

Intercept	*SIZE*	*RETURN*	Beta	R^2_{adjusted}
	Coefficient Estimates			
	Panel A: Annual Percentage Returns			
14.443	−0.364	−0.254	5.438	0.68
(10.517)	(−3.779)	(−2.996)	(1.707)	
	Panel B: Monthly Percentage Returns, All Months			
1.236	−0.031	−0.023	0.369	0.68
(4.671)	(−2.926)	(−3.039)	(1.393)	

$$r_p - r_f = a_0 + a_1 SIZE_p + a_2 DS \cdot RETURN_p + a_3 DM \cdot RETURN_p + a_4 DL \cdot RETURN_p + a_5 Beta_p + e_p$$

Table 4 (*Continued*)

Intercept	*SIZE*	*DS* · *RETURN*	*DM* · *RETURN*	*DL* · *RETURN*	Beta	R^2_{adjusted}
		Coefficient Estimates				
		Panel C: Annual Percentage Returns				
18.113	−0.597	−0.417	−0.182	−0.136	4.364	0.72
(9.915)	(−5.440)	(−4.257)	(−2.009)	(−1.433)	(1.298)	
		Panel D: Monthly Percentage Returns, All Months				
1.631	−0.055	−0.039	−0.018	−0.010	0.238	0.73
(6.431)	(−5.675)	(−4.733)	(−2.235)	(−1.326)	(0.898)	

The empirical results displayed in Table 4, Panel A, show the following:

- The return on the portfolio …
  - … decreases with firm size
  - … decreases with the ranking-period raw return
  - … increases with the market beta.
- The *RETURN* coefficient of -0.254 implies that after controlling for size and systematic risk (beta), extreme losers outperform extreme winners by an average 4.8 percentage points per annum for the five post-ranking years
- The coefficient on beta of 5.438 percent is lower than the 9.5 percent reported in Figure 1a, which means that the empirical securities market line is even flatter than the one shown in Figure 1a
  - The results displayed in Figure 1a might suffer from an omitted variable bias, because the analysis did not control for firm size
  - Quantitatively, the overreaction effect is nearly as great as the small firm effect.

Panel B of Table 4 shows results from monthly returns (rather than annual returns

- The results, after multiplying by 12, are quantitatively similar to the those in Panel A
  - The overreaction effect is slightly stronger; extreme losers outperform extreme winners in the five-year post-ranking period by an annual 5.2 percentage points
  - The compensation per unit of beta (systematic risk) amounts to 4.4 percent per year, which falls short of the 5.4 percent reported in Panel A for annual return data.

In Panels C and D, the authors allow the overreaction effect to vary by firm size (small caps, mid caps, and large caps)

- The overreaction is strongest for small caps, much weaker for mid caps and statistically insignificant for large caps
  - Among small caps, extreme losers outperforms extreme winners in the five-year post-ranking period by a staggering 7.9 percentage points per year
    - For mid caps, this difference is only 3.5 percentage points, while for large caps it is a statistically insignificant 2.6 percentage points.
- Note that the regression coefficient of beta is insignificant in Panel C (annual returns) and in Panel D (monthly returns)
  - The regression coefficient of beta is insignificant in Panel B (monthly returns) and barely significant in Panel A (annual returns).

In Table 5, the authors elaborate further on the influence of firm size on overreaction

For each <u>size decile</u>, the authors offer estimates for the following regression equation:

$$r_p - r_f = a_0 + a_1 \cdot RETURN_p + a_2 \cdot Beta_p + e_p \ , \ E[e_p] = 0$$

Each of the 10 size deciles uses 40 portfolios out of the 400 cells formed for the analysis displayed in Table 4

Size decile 1 comprises the smallest, and size decile 10 comprises the largest companies.

Table 5 OLS regressions of annual average percentage excess returns on ranking-period returns and beta by size decile.

RETURN is measured 1 to 20 (1 = losers, 20 = winners), where prior returns are measured over the five years prior to the portfolio formation date. Firms are assigned to size deciles (1 = small, 10 = large) on the basis of their market capitalization at the end of the ranking period. The beta of each portfolio is calculated as the pooled (over firms and post-ranking years) beta. Each of the ten regressions uses forty observations (two ranks of size with twenty prior-return portfolios in each size rank). *T*-statistics, computed using the fourth-order autoregressive process described in table 4, are in parentheses.

$$r_p - r_f = a_0 + a_1 RETURN_p + a_2 Beta_p + e_p$$

	Coefficient Estimates				
Size Decile	Intercept	*RETURN*	Beta	R^2_{adjusted}	$-19 \times RETURN$ Coefficient[a]
1	9.888 (2.463)	−0.578 (−2.119)	9.980 (2.670)	0.76	10.98%
2	27.658 (4.379)	−0.729 (−6.436)	−2.784 (−0.426)	0.74	13.85%
3	21.218 (4.723)	−0.510 (−3.382)	0.402 (0.078)	0.65	9.69%
4	18.942 (6.730)	−0.350 (−3.811)	0.739 (0.242)	0.51	6.65%
5	16.356 (3.715)	−0.140 (−2.629)	−0.641 (−0.101)	0.10	2.66%
6	14.226 (1.982)	−0.293 (−2.242)	2.489 (0.288)	0.52	5.57%
7	9.149 (4.691)	−0.153 (−1.755)	4.838 (2.463)	0.51	2.91%
8	8.018 (3.012)	−0.113 (−0.764)	5.171 (1.000)	0.37	2.15%
9	6.101 (1.634)	−0.016 (−0.149)	4.524 (0.572)	0.01	0.30%
10	5.080 (1.932)	0.040 (0.327)	2.471 (0.466)	0.01	−0.76%

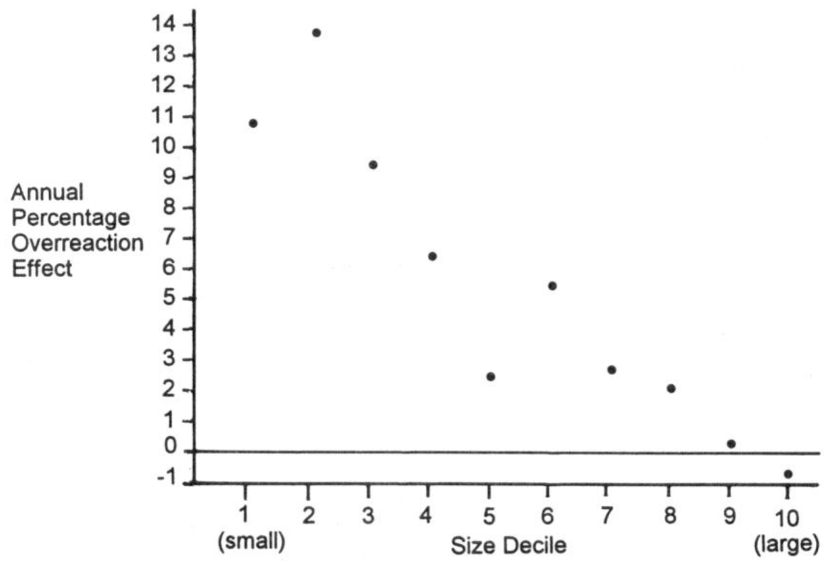
a. Multiplying the coefficients on *RETURN* by −19 gives the expected difference in annual returns for the five post-ranking years between prior-return portfolios 1 and 20, controlling for beta, for firms categorized by their size decile.

Table 5 shows that the absolute value of the regression coefficient of the *RETURN* variable decreases with firm size, which indicates that the overreaction effect declines with firm size

The last column of Table 5 shows the implied annual percentage point difference between the extreme loser and the extreme winner portfolios, holding size and beta constant

The differences are plotted in Figure 4

Figure 4 The difference in annual abnormal returns between extreme loser and winner portfolios by size decile.



The numbers plotted are the coefficients on *RETURN* in table 5 multiplied by −19. This represents the expected difference in annual returns for the five post-ranking years between prior return portfolios 1 and 20, controlling for beta, for firms categorized by their size decile.

Figure 4 demonstrates an overreaction effect for small caps of 10 percentage points per year (which results in a staggering 50 percent per five years, even before compounding)

- For the largest 20 percent of NYSE companies (roughly the S&P 500), there is no evidence of overreaction
- Because small caps are primarily owned by individuals while large caps are primarily owned by institutional investors, the results indicate that while individuals overreact, institutions do not
- The regression coefficient for beta varies greatly across size deciles; it is statistically significant only for two of the ten size deciles
  - For the largest two size deciles, which constitute the bulk of the market capitalization, beta is far from statistically significant.
- Also, for the largest two deciles the values of the R^2 statistics are virtually zero
  - With no evidence for market sensitivity or overreaction among large caps, the authors conclude that, for large caps, a stock is a stock.

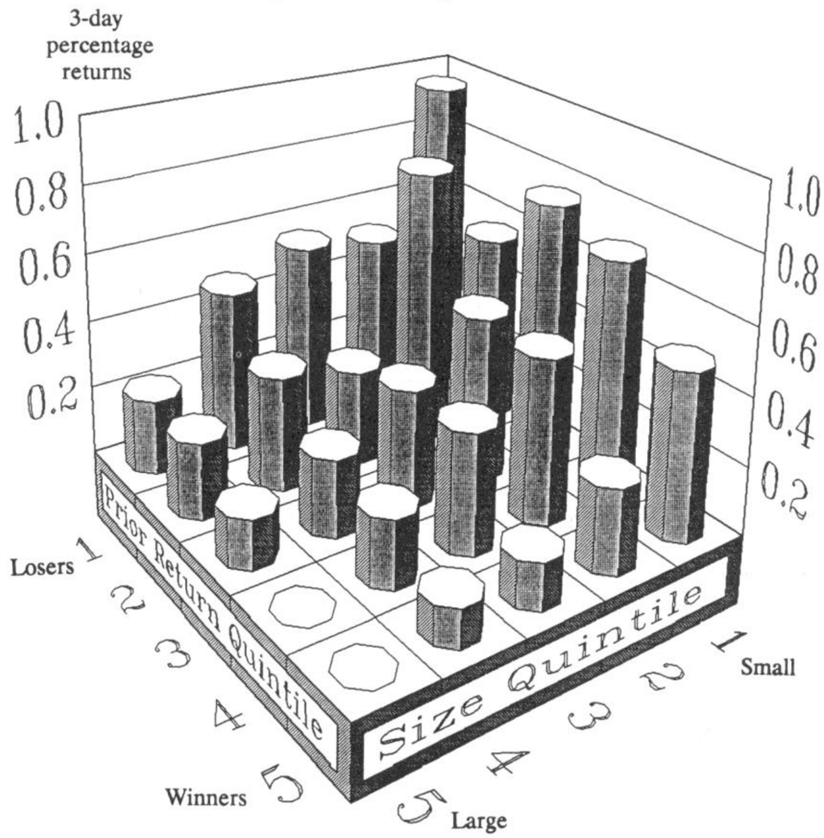
Evidence for overreaction to earnings announcements

The authors establish further evidence of overreaction by analyzing stock returns around earnings announcements

The authors restrict themselves to (five-year) ranking periods that end in the period 1970-81, for which they find 227,522 earnings announcements on file

For each earnings announcement, the authors calculate the raw return of the respective stock in a three-day window, the last day of which is the day of the announcement

Figure 5 The joint distribution of three-day earnings announcement returns categorized by market capitalization and prior returns.



Firms are assigned to portfolios based upon independent rankings of size and prior returns. The average three-day raw return at subsequent earnings announcements is computed for Compustat-listed quarterly earnings announcement dates during the five-year post-ranking period. The average three-day raw return is 0.001% for the largest extreme winners and 0.958% for the smallest extreme losers.

Figure 5 plots the three-day earnings announcement period returns using the same firm-size quintiles and ranking-period return quintiles used in Figures 2 and 3

- The small losers have average three-day returns of 0.958 percent, while the large winners have three-day returns of 0.001 percent
- Returning to the twenty portfolios (used in the prior quantitative analyses), the average earnings announcement return for extreme losers is 0.63 percent, while the corresponding return for the extreme winners is zero
  - Thus, the evidence from earnings announcements indicates that the market is systematically surprised at subsequent earnings announcements in a manner consistent with the overreaction hypothesis.

Is the earnings announcement effect displayed in Figure 5 due to firm size, firm risk or even both?

- V.V. Chari, R. Jagannathan, and A.R. Ofer (1988, "Seasonalities in Security Returns: The Case of Earnings Announcements," *Journal of Financial Economics* 21, 101-121) document that small caps have higher earnings-announcement-period returns than large caps
- Also, Chari, Jagannathan, and Ofer hypothesize that because of the increased flow of information around earnings announcements, these periods are riskier than non-announcement periods

Chopra, Lakonishok and Ritter thus carry out a multivariate regression analysis that allows them to control for firm size and systematic risk (beta) simultaneously

For each of the ranking periods that end in the period 1970-1981, firms are, independently, …

… ranked into 20 size categories by market value at the end of the five-year ranking period

… ranked into 20 return categories based on the five-year ranking period (raw) return.

There are 400 cells, which might not all be of the same number of firms n_k , $k = 1,...,400$

Six cells are dropped because they have fewer than 100 observations (earnings announcements).

In the <u>first step</u>, for each of the 394 cells (portfolios, $p = 1,...394$), <u>for the three-day earnings-announcement window</u> the market beta is estimated from a pooled (across post-ranking years, $\tau = 1,...,5$, and companies, $j = 1,...,n_k$) regression of the following <u>market model</u>:

$$r_{j,\tau} = \alpha + \beta \cdot r_m + \varepsilon, \; E[\varepsilon] = 0$$

where r_m is the respective three-day market return.

In the <u>second step</u>, the portfolio's excess return is regressed on the portfolio's size category, ranking-period raw return category, and beta:

$$r_p = a_0 + a_1 \cdot SIZE_p + a_2 \cdot RETURN_p$$
$$+ a_3 \cdot Beta_p + e_p \; , \; E[e_p] = 0$$

Table 6 Regression of three-day earnings announcement portfolio returns on size, prior returns, and beta.

394 portfolios are used (400 portfolios based on independently ranking firms by size and prior return, with six portfolios deleted which had fewer than 100 earnings announcements). Size is measured with the smallest firms in portfolio 1, and the largest in portfolio 20. Prior returns (measured over the five prior years) are also ranked from 1 to 20, with 1 being the losers. Betas are calculated for each portfolio using all earnings announcement returns for all firms in the portfolio. The dependent variable is measured as the percentage return per three-day announcement period [−2, 0], for earnings announcements made during the first five post-ranking years. Earnings announcement days are from Compustat's industrial, historical, and research tapes, for announcements during the five post-ranking years following the ranking periods ending in 1970–81. There are 227,522 earnings announcements. *T*-statistics, computed using the time-series variance of the cross-sectional regression coefficients, adjusted for first-order autocorrelation, are in parentheses.

$$R_p = a_0 + a_1 SIZE_p + a_2 RETURN_p + a_3 Beta_p + e_i$$

Coefficient Estimates				
Intercept	*SIZE*	*RETURN*	Beta	R^2_{adjusted}
0.641 (3.230)	−0.027 (−7.701)	−0.014 (−2.548)	0.111 (2.018)	0.32

The regression coefficient shown in Table 6 indicate that, holding beta and firm size constant, the earnings announcement returns are greater for prior ranking-period losers than for prior winners

- Multiplying the coefficient of -0.0142 by -19 results in an abnormal return of the extreme loser portfolio over the extreme winner portfolio of 0.27 percentage points per announcement
  - Since there are four (quarterly) earnings announcements per year, the extreme loser portfolio outperforms the extreme winner portfolio by 1.08 percentage points during these 12 trading days alone.

Conclusion

There is overreaction, momentum, and mean reversion in the stock market, consistent with the evidence provided by De Bondt and Thaler (1986)

Due to overreaction, extreme loser portfolios outperform extreme winner portfolios by (depending on the approach employed) an annual 5-10 percentage points in the five years following the five-year ranking period

The evidence for overreaction is buttressed by findings of abnormal returns due to stock price movements around quarterly earnings announcements

Overreaction is stronger for small caps, which are predominantly held by individuals; there is only weak evidence for overreaction in large caps, which are predominantly held by institutional investors

Institutions—by virtue of having an infinite time horizon—might prevent overreaction by picking up stocks that are dumped by individuals in response to bad news.

There is a strong seasonal return pattern in mean reversion

The return reversal (abnormal positive returns for portfolios based on five-year formation periods) is entirely due to the month of January (Table 3, columns (4-6), row "$r_1 - r_{20}$").

There is momentum, which indicates that overreaction takes time to run its course

- Portfolios based on one-year formation periods show negative abnormal returns over the year following the portfolio formation period (Table 3, columns (10-12), row "$r_1 - r_{20}$ in year $+1$")
- The return reversal that has taken effect after five years is smaller for the one-year formation period (due to momentum) than for the five-year formation period as momentum has fizzled out (in Table 3, row "$r_1 - r_{20}$," compare columns (10-12) to columns (4-6)).

Investment advice

- For large caps (e.g., the S&P 500), there is only weak evidence for overreaction (no evidence in Fig. 4; weak evidence in Fig. 5 [earnings announcements])
  - The deep value approach (picking stocks that are trading at around 60 percent of intrinsic value) is more promising for mid caps and small caps than for large caps
  - The further one moves away from large caps, the more likely it is that stocks take prolonged swings away from their fundamental value
    - Small cap portfolios might lag the performance of large cap portfolios for several years before the return reversal in the small cap portfolio sets in.

[Blank page]

Empirical Studies

13. The Long-Run Performance of IPOs

Reference:

Bodie, Zvi, Alex Kane, and Alan J. Marcus (1999) *Investments*. 4th ed. Boston: Irwin—McGraw-Hill.

Ibbotson, Roger G., and Jay R. Ritter (1995) "Initial Public Offerings," in: Robert A. Jarrow, Vojislav Maksimovic, and William T. Ziemba, ed., *Finance*, Handbooks in Operations Research and Management Science, Vol. 9, Chap. 30, Amsterdam: Elsevier, 993-1016.

Ritter, Jay R. (1991) "The Long-Run Performance of Initial Public Offerings," *Journal of Finance* 46, 3-27, reprinted in: Richard H. Thaler, ed. (1993) *Advances in Behavioral Finance*. New York: Russell Sage Foundation, 459-489.

Ritter, Jay R., and Ivo Welch (2002) "A Review of IPO Activity, Pricing, and Allocations," *Journal of Finance* 57, 1795-1828.

Terminology and administration

One way of raising equity capital is to sell shares to the public

The market for shares that are yet to be issued is called the primary market

After issuance, shares trade in the aftermarket—a term commonly used in conjunction with IPOs—or, synonymously, the secondary market.

There are two types of public share offerings

In initial public offerings (IPOs), stocks are issued by a heretofore privately owned company

In seasoned equity offerings, traded companies issue additional stock.

In private placements, stocks are issued not to the public but to a few institutional investors or wealthy individuals

Public offerings of securities are typically marketed through underwriting by investment banks

There is a lead underwriter, which—along with the other investment banks—forms the underwriting syndicate

The investment banks advise the issuing firm regarding the terms on which it should market the securities.

There are two types of underwriting procedures

In a firm-commitment underwriting agreement—the most common type of underwriting in the United States—investment banks purchase the securities from the issuing company and resell them to the public less a spread—typically seven percent—that serves as compensation for the underwriters

In a best-effort underwriting agreement the investment banks agree to help the firm sell the issue to the public but do not actually purchase the securities.

Public offerings are regulated

A preliminary registration statement must be filed with the SEC (Securities and Exchange Commission; http://www.sec.gov), describing the equity issue and the prospects of the company

This preliminary prospectus is known as a red herring because of a statement printed in red ink stating that the company shall not attempt to sell the security before the registration is approved

After the SEC has approved the registration statement filed by the issuer, it is called a prospectus

At this time, the investment bank announces to the public an offer price—when there is fixed-price underwriting—or a price range—when there is bookbuilding.

Bookbuilding

Once the SEC has commented on the registration statement and a preliminary prospectus has been distributed to prospective investors, the investment banks organize a road show in which the issuer and the underwriter market the company

Large investors communicate their interest in purchasing share of the IPO to the underwriters

These indications are called books, and the process of pooling potential investors is called bookbuilding

It is common for an investment bank to revise, based on feedback from the investing community, both the initial estimate of the offer prices—that is, the midpoint of the announced price range—and the number of shares to be offered

Shares of IPOs are allocated to investors in part based on the strength of each investor's expressed interest.

Underwriters do not only have price discretion, they also have quantity discretion

- In allocating shares, they do not only control who gets shares, they also decide how many shares are allocated in total
  - Almost all IPOs contain an overallotment option—called Green Shoe option after the first company that included it in its 1963 IPO—for up to 15 percent of the shares offered
    - Then, if the price weakens the underwriter can—but does not have to—buy back the extra 15 percent and retire the shares as if they had not been issued in the first place
    - The Green Shoe serves as an insurance of the investment bank against aggravating clients through rationing, increasing the bank's marketing effort.

There are <u>three anomalies</u> in the pricing of initial public stock offerings

First, there is short-run underpricing

Starting with an SEC report in 1963, numerous studies have provided evidence for positive average <u>initial returns</u> of IPOs

Typically, initial returns are calculated by comparing the offer price to the market price at the end of the first day of trading.

Second, there are "hot issue" markets—that is, periods of extraordinary high volume of IPO activity

During hot issue markets, IPO underpricing is considerably more pronounced than when IPO business is slow.

Third, there is long-run underperformance

An ongoing discussion about methodological issues in measuring long-run IPO performance notwithstanding, it is fair to say that there is solid evidence of IPOs underperforming the market—or a portfolio of comparable companies, for that matter—in the long haul.

Overview on underpricing, volume, and long-run performance

Underpricing is universal, although the extent of underpricing seems to vary across countries

Table 1

International evidence on short-run IPO underpricing

Country	Sample size	Time period	Average initial return (%)
Australia	266	1976–89	11.9
Belgium	28	1984–90	10.2
Brazil	62	1979–90	78.5
Canada	258	1971–92	5.4
Chile	19	1982–90	16.3
Finland	85	1984–92	9.6
France	187	1983–92	4.2
Germany	172	1978–92	11.1
Greece	79	1987–91	48.5
Hong Kong	80	1980–90	17.6
India	98	1992–93	35.3
Italy	75	1985–91	27.1
Japan	472	1970–91	32.5
Korea	347	1980–90	78.1
Malaysia	132	1980–91	80.3
Mexico	37	1987–90	33.0
Netherlands	72	1982–91	7.2
New Zealand	149	1979–87	28.8
Portugal	62	1986–87	54.4
Singapore	128	1973–92	31.4
Spain	71	1985–90	35.0
Sweden	213	1970–91	39.0
Switzerland	42	1983–89	35.8
Taiwan	168	1971–90	45.0
Thailand	32	1988–89	58.1
United Kingdom	2,133	1959–90	12.0
United States	10,626	1960–92	15.3

Initial returns are generally defined as the percentage increase from the offer price to a closing market price shortly after public trading begins. The length of this period varies from study to study, with one day to several weeks being the usual time frame. The studies using only a day or two generally report raw returns, whereas those using a week or more generally adjust for market movements during the measurement interval. The average initial returns are generally not very sensitive to the length of the interval used or whether market movements are accounted for, since all of the measurement intervals are short. The average initial returns are computed as equally-weighted averages of the initial returns. The data are from various studies, including those by Lee, Taylor & Walter [1994a] for Australia; Aggarwal, Leal & Hernandez [1993] for Brazil, Chile, and Mexico; Jog & Srivastava [1995] for Canada; Dawson [1987] for Hong Kong; Dimson [1979] and Levis [1993] for the U.K.; Husson & Jacquillat [1989] and Leleux & Muzyka [1993] for France; Ibbotson, Sindelar & Ritter [1994] for the U.S.; Jenkinson [1990], and Hebner & Hiraki [1993] for Japan; Kazantzis & Levis [1994] for Greece; Keloharju [1993] for Finland; Kim, Krinsky & Lee [1993] for Korea; Krishnamurti & Kumar [1994] for India; Isa [1993] for Malaysia; Lee, Taylor & Walter [1994b] for Singapore; Rydqvist [1993] for Sweden; Kunz & Aggarwal [1994] for Switzerland; Uhlir [1989] and Ljungqvist [1993] for Germany; Vos & Cheung [1992] for New Zealand; Chen [1992] for Taiwan; Wethyavivorn & Koo-smith [1991] for Thailand; and Wessels [1989] for the Netherlands. Additional references are listed in Loughran, Ritter & Rydqvist [1994].

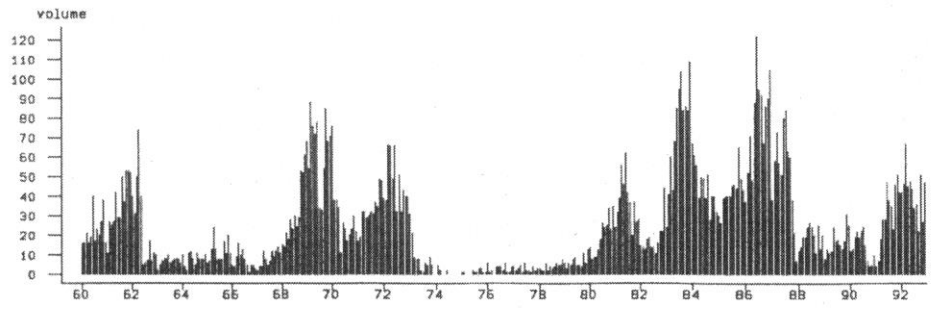
Source: Ibbotson, Roger G., and Jay R. Ritter (1995) "Initial Public Offerings," in: Robert A. Jarrow, Vojislav Maksimovic, and William T. Ziemba, ed., *Finance*, Handbooks in Operations Research and Management Science, Vol. 9, Ch. 30, Amsterdam: Elsevier, 993-1016.

A note of caution on the calculation of initial returns

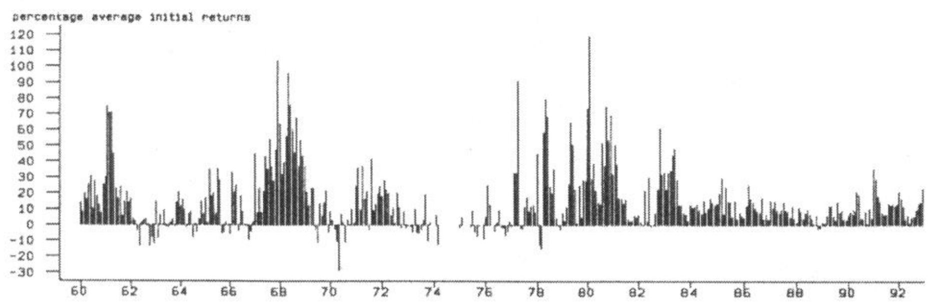
If initial returns are calculated simply by averaging across IPOs, disregarding size differences across IPOs, underpricing might be overestimated if small IPOs are more underpriced than large IPOs

Ideally, one might want to calculate underpricing from a <u>weighted</u> average of long-run IPO returns.

Underpricing figures prominently when IPO activity is strong—that is, in hot issue markets



The number of IPOs by month in the U.S. during 1960–1992, excluding closed-end fund IPOs. Source: Ibbotson, Sindelar & Ritter [1994].



Average initial returns by month for S.E.C.-registered IPOs in the U.S. during 1960–1992. Source: Ibbotson, Sindelar & Ritter [1994].

Source: Ibbotson, Roger G., and Jay R. Ritter (1995) "Initial Public Offerings," in: Robert A. Jarrow, Vojislav Maksimovic, and William T. Ziemba, ed., *Finance*, Handbooks in Operations Research and Management Science, Vol. 9, Ch. 30, Amsterdam: Elsevier, 993-1016.

IPOs do not always underperform the market in the long haul—yet, stocks IPO'ed during hot issue markets tend to perform miserably

Number of IPOs, First-day Returns, Gross Proceeds, Amount of Money Left on the Table, and Long-run Performance, by Cohort Year, 1980 to 2001

The equally weighted (EW) average first-day return is measured from the offer price to the first CRSP-listed closing price. Gross proceeds is the amount raised from investors in millions (2001 purchasing power using the CPI, global offering amount, excluding overallotment options). Money left on the table (millions of dollars, 2001 purchasing power) is calculated as the number of shares issued times the change from the offer price to the first-day closing price. EW average three-year buy-and-hold percentage returns (capital gains plus dividends) are calculated from the first closing market price to the earlier of the three-year anniversary price, the delisting price, or September 30, 2001. Buy-and-hold returns for initial public offerings (IPOs) occurring after September 30, 2000 are not calculated. Market-adjusted returns are calculated as the buy-and-hold return on an IPO minus the compounded daily return on the CRSP value-weighted index of AMEX, Nasdaq, and NYSE firms. Style-adjusted buy-and-hold returns are calculated as the difference between the return on an IPO and a style-matched firm. For each IPO, a non-IPO matching firm that has been CRSP listed for at least five years with the closest market capitalization and book-to-market ratio as the IPO is used. If this is delisted prior to the IPO return's ending date, or if it conducts a follow-on stock offering, a replacement matching firm is spliced in on a point-forward basis. IPOs with an offer price below $5.00 per share, unit offers, REITs, closed-end funds, banks and S&Ls, ADRs, and IPOs not listed on CRSP within six months of issuing have been excluded. Data is from Thomson Financial Securities Data, with supplements from Dealogic and other sources, and corrections by the authors.

					Average 3-Year Buy-and-Hold Return		
Year	Number of IPOs	Average First-day Return	Aggregate Gross Proceeds, Millions	Aggregate Money Left on the Table, Millions	IPOs	Market-Adjusted	Style-Adjusted
1980	70	14.5%	$ 2,020	$ 408	88.2%	35.5%	17.1%
1981	191	5.9%	$ 4,613	$ 264	12.8%	−26.2%	−7.4%
1982	77	11.4%	$ 1,839	$ 245	32.2%	−36.5%	−48.7%
1983	442	10.1%	$ 15,348	$ 1,479	15.4%	−38.7%	2.5%
1984	172	3.6%	$ 3,543	$ 86	27.7%	−51.3%	3.0%
1985	179	6.3%	$ 6,963	$ 354	7.6%	−39.5%	7.3%
1986	378	6.3%	$ 19,653	$ 1,030	18.6%	−20.4%	14.3%
1987	271	6.0%	$ 16,299	$ 1,019	−1.8%	−18.9%	4.5%
1988	97	5.4%	$ 5,324	$ 186	55.7%	8.3%	51.3%
1989	105	8.1%	$ 6,773	$ 336	51.1%	16.8%	32.5%
1990	104	10.8%	$ 5,611	$ 454	12.2%	−34.1%	−32.4%
1991	273	12.1%	$ 15,923	$ 1,788	31.5%	−1.7%	5.8%
1992	385	10.2%	$ 26,373	$ 2,148	34.8%	−2.3%	−19.4%
1993	483	12.8%	$ 34,422	$ 3,915	44.9%	−7.8%	−23.9%
1994	387	9.8%	$ 19,323	$ 1,650	74.1%	−8.3%	1.0%
1995	432	21.5%	$ 28,347	$ 5,033	24.8%	−62.3%	−14.1%
1996	621	16.7%	$ 45,940	$ 7,383	25.6%	−57.0%	8.6%
1997	432	13.8%	$ 31,701	$ 4,664	67.7%	6.8%	41.0%
1998	267	22.3%	$ 34,628	$ 5,352	27.1%	9.1%	12.2%
1999	457	71.7%	$ 66,770	$ 37,943	−46.2%	−32.9%	−74.2%
2000	346	56.1%	$ 62,593	$ 27,682	−64.7%	−36.4%	−42.6%
2001	80	14.0%	$ 34,344	$ 2,973	n.a.	n.a.	n.a.
1980–1989	1,982	7.4%	$ 82,476	$ 5,409	20.8%	−24.7%	6.9%
1990–1994	1,632	11.2%	$101,652	$ 9,954	44.7%	−7.2%	−12.7%
1995–1998	1,752	18.1%	$140,613	$ 22,436	36.0%	−32.3%	11.6%
1999–2000	803	65.0%	$129,363	$ 65,625	−53.8%	−34.3%	−61.2%
2001	80	14.0%	$ 34,344	$ 2,973	n.a.	n.a.	n.a.
1980–2001	6,249	18.8%	$488,448	$106,397	22.6%	−23.4%	−5.1%

Source: Ritter and Welch (2002).

The period 1980-2001

- The number of IPOs per year varied between less than 100 and more than 400
  - These IPOs raised $488 billion (in 2001 dollars) in gross proceeds—an average $78 million per deal
  - At the end of the first day of trading, the shares of these IPOs traded an average 18.8 percent above the price at which the company sold them
  - For an investor, buying shares at the first-day closing price and holding them for three years, IPOs returned 22.6 percent
    - Still, over three years, the average IPO underperformed seasoned companies with the same market capitalization and book-to-market ratio by 5.1 <u>percentage points</u>.

The 1980-2001 averages hide a great deal of year-by-year variation

The 1980s saw modest IPO activity (volume of about $8 billion per year); in the 1990s, volume roughly doubled to $20 billion per year during the period 1990-1994, then doubled again from 1995 to 1998 ($35 billion per year), before doubling again from 1999 to 2000 ($65 billion per year); in 2000, volume fell to $34 billion

Average first-day returns show a similar pattern, increasing from 7.4 percent in the 1980s to 11.2 percent in the early 1990s, to 18.1 percent in the mid-1990s, and to 65.0 percent in the period 1999-2000—the peak of the bubble—before dropping to 14.0 percent in 2001

The long-run performance of IPOs also varies over time; three-year market-adjusted buy-and-hold returns are negative in each of the aforementioned subperiods (see table above) but not for every cohort year; style-adjusted buy-and-hold returns are not as reliably negative, with many cohorts and some subperiods having positive style-adjusted buy-and-hold returns.

Why do firms go public?

In most cases, the primary answer is to raise equity capital for the company and to create a public market in which the founders and other shareholders (e.g., venture capitalists) can convert some of their equity into cash at a future date

Absent cash considerations, most entrepreneurs would rather run their companies than concern themselves with the complex public market process.

Then again, why is the motivation to do an IPO stronger in some time periods than in others?

For instance, in 1969, there were 683 IPOs, more than in the period 1935 through 1959 taken together.

Hypotheses of underpricing

It is fair to say that there is no single dominant cause for underpricing

It is not so much a matter of which model is right but a matter of the relative importance of different models

Furthermore, one reason can be of more importance for some companies or at some times.

Hypotheses of going public are difficult to test

First, there is a problem of observational equivalence—the fact that a given outcome may be compatible with a variety of data generating processes

For instance, identical prices at the pump—which implies simultaneous price hikes—are compatible with perfect competition and with collusion.

Second, we know of those companies only that go public but not of those would have gone public under different circumstances

A study who interviewed death row inmates on the deterring effect of the death penalty concluded that capital punishment does not prevent capital crime

The study ignored those who have possibly been deterred from committing a capital crime and, hence, are not on death row.

Life cycle hypotheses

- By going public, entrepreneurs help facilitate the acquisition of their company for a higher value than what they would get from an outright sale
  - This is because it is easier for a potential acquirer to spot a target that is public rather than private
  - Moreover, it is said that acquirers can pressure targets on pricing concessions more than they can pressure outside investors
  - Then again, it has been found that entrepreneurs often regain control from the venture capitalist in venture-capital-backed companies at the IPO, indicating that the IPO was more an exit strategy to the venture capitalist than to the entrepreneur.
- Another hypothesis states that companies go public when the marginal benefit of doing so outweighs the marginal cost
  - Pre-IPO investors (e.g., venture capitalists) are endowed with control rights but are undiversified—the lack of diversification causing a discount in firm valuation
  - As the problem of non-verifiable cash flows eases with an increase in public information about the company, the advantage of having a concentrated shareholder structure lessens.

Market timing hypotheses

- If the stock market places too low a valuation on the firm, the entrepreneur, who (thinks he) knows the intrinsic value of the firm, will wait with the IPO till the market offers more favorable prices, making use of windows of opportunity
  - The cynical version of this hypothesis is that in bull markets, people have an incentive of creating companies, of which they know are garbage, just to take them public.
- There is strong empirical evidence that entrepreneurs make use of windows of opportunity in the IPO market
  - Then again, why don't investors recognize this or, in other words, how can hot issue markets with their dismal long-run IPO performance repeat themselves?

The changing composition of IPO issuers

Aggregate numbers disguise the fact that the characteristics of companies going public have changed over the years

It has been shown that the median age of companies going public was stable at around 7 years since 1980 till the period 1999-2000, when the median age dropped to 5 years before climbing to 12 years in 2001

The table above shows that the percentage of technology stocks increased sharply in the late 1990s before dropping in 2001

(Tech stocks are defined as Internet stocks, computer software and hardware, communications equipment, electronics, navigation equipment, measuring and controlling devices, medical instruments, telephone equipment, and communications services; biotechnology is not included.)

Fraction of IPOs with Negative Earnings (Trailing Last 12 Months), 1980 to 2001

IPOs with an offer price below $5.00 per share, unit offers, ADRs, closed-end funds, REITs, bank and S&L IPOs, and firms not listed on CRSP within six months of the offer date are excluded. When available, we use the earnings per share for the most recent 12 months (commonly known as LTM for last 12 months) prior to going public. When a merger is involved, we use the pro forma numbers (as if the merger had already occurred). When unavailable, we use the most recent fiscal year EPS numbers. Missing numbers are supplemented by direct inspection of prospectuses on EDGAR, and EPS information from Dealogic (also known as ComScan) for IPOs after 1991, and Howard and Co.'s *Going Public: The IPO Reporter* from 1980 to 1985. Tech stocks are defined as Internet-related stocks plus other technology stocks, not including biotech. Loughran and Ritter (2001) list the SIC codes in their appendix 4.

Time Period	Number of IPOs	Percentage Tech Stocks	Percentage of IPOs with EPS < 0	Mean First-day Returns	
				EPS < 0	EPS ≥ 0
1980–1989	1,982	26%	19%	9.1%	6.8%
1990–1994	1,632	23%	26%	10.8%	11.4%
1995–1998	1,752	37%	37%	19.2%	17.4%
1999–2000	803	72%	79%	72.0%	43.5%
2001	80	29%	49%	13.4%	14.6%
1980–2001	6,249	34.5%	34%	31.4%	12.5%

Source: Ritter and Welch (2002).

Remarkably, the number of non-tech stock IPOs remained at a level of about 100 per year before, during, and after the bubble (not shown)

In other words, the fluctuation in the number of IPOs was entirely due to tech stocks.

The increase in the percentage of technology firms over time is mirrored in the number of firms with negative earnings in the 12 months prior to going public

- Note that is was unusual for a prestigious investment bank in the 1960s and 1970s to take a company public that did not have at least four years of positive earnings
- In the 1980s, four quarters of positive earnings was still standard whereas in the 1990s fewer and fewer companies met this threshold—still, the investment bank's analysts would project profitability in the year after going public
- Then, during the period 1999-2000—the peak of the bubble—companies with no immediate prospect of becoming profitable became common
  - For instance, public forecasts for eToys, which liquidated in 2001, projected no profits for at least two years—the EPS (earnings per share) forecasts were -\$0.27 for 1999 and -\$0.55 for 2000.
- The table above shows that, over time, both the fraction of negative pre-IPO EPS companies and the degree of underpricing increased
  - On the other hand, only in the period 1999-2000—the peak of the bubble—there is a cross-sectional relation between the sign of the pre-IPO earnings and underpricing.

Hypotheses of short-run underpricing

Hypotheses based on asymmetric information

If the issuer—typically, the entrepreneur—is more informed than the investors, rational investors fear a lemon problem:

Only issuers with worse than average quality are willing to sell their shares at the average price

To distinguish themselves from the pool of low-quality issuers, high-quality issuers may attempt to signal their quality.

In signaling models of IPO underpricing, better-quality issuers deliberately sell their shares at a lower price than the market believes they are worth, which deters lower quality issuers from imitating—in other words, companies voluntarily leave money on the table

With some patience—so these models claim—these issuers can recoup their up-front sacrifice post-IPO, either in future issuing activity, favorable market responses to future dividend announcements, or analyst coverage

Unfortunately, the empirical evidence on these signaling models is mixed, at best

From personal experience I know that signaling works only if the other party is educated enough to understand the signaling model—a proposition that, in my view, assumes far too much.

If the investors are more informed than the issuer, then the issuer faces a placement problem

For instance, the issuer might not know the demand for its stock

One way to model this problem is to assume that all investors are equally informed

Then, only IPOs that are priced fair (that is, priced at intrinsic value) or are underpriced, will succeed

Consequently, we would observe many underpriced and no overpriced IPOs

Then again, there are some overpriced IPOs being observed in the real word, which may either due to (unbiased) errors in investors' expectations or due to the fact that not all investors are equally informed.

More realistic than equally informed investors is the assumption that investors differ in the degree of information they have

- In this case, pricing too high might induce investors and issuers to fear a winner's curse or a negative cascade
  - In a winner's curse situation, investors fear that they will only receive full allocations if they happen to be among the most optimistic investors—something they don't know—while when everyone desired the IPO, they get rationed
    - Consequently, an investor would receive a full allocation of overpriced IPOs but only a partial allocation of underpriced IPOs
      - Thus, the investor's average return, conditioned on receiving shares, would be below the unconditional return, which implies that IPOs have to be underpriced.
    - An empirical study on IPOs in Singapore that uses information on rationing shows that an uninformed strategy just breaks even.
  - In an informational cascade, investors attempt to judge the interest of other investors
    - Investors only request shares if they believe that the offering is hot
      - Thus, pricing just a little too high leaves the issuer with a probability of failure—a situation where investors abstain because other investors abstain.
    - In support of this hypothesis it has been found that IPOs tend to be either undersubscribed or hugely oversubscribed, only a few being mildly oversubsribed.

Bookbuilding theories

- The common practice of bookbuilding allows underwriters to obtain information from informed investors
  - The road show helps underwriters to gauge demand as they record indications of interest from potential investors
    - If there is strong demand, the underwriter will notch up the offer price
      - But if potential investors know that showing a high willingness to pay will result in a higher offer price, these investors must be offered something in return to give them an incentive to truthfully reveal their preferences.
- Studies have found that underwriters do not fully adjust the offer price upward when demand is strong
  - Thus, when underwriters revise the share price upward from their original estimate in the preliminary prospectus, underpricing tends to be higher.

Hypotheses based on *symmetric* information

It has been argued that issuers underprice to insulate them against law suits of disgruntled shareholders should the IPO'ed company perform badly

Clearly, all else equal, the lower the offer price, the higher the—possibly negative—long-run IPO return

The most convincing evidence against this hypothesis is that underpricing is observed in countries in which the legal system does not allow for U.S. style litigation.

Another hypothesis states that investment banks are leaning against the wind of excessive investor optimism

The argument is unconvincing because investment banks engaged in a series of activities that encouraged overvaluations, among them issuing "buy" and "strong buy" recommendations at the end of the 25-calendar-day post-IPO quiet period even when market prices had risen far above the offer price

For instance, Credit Suisse First Boston (CSFB) took Corvis public on July 28, 2000, at an offer price of $36.00. At the closing price of $84.719 on the first day of trading, the first-day return was 135 percent. When the quiet period ended, the five co-managing underwriters all put out "buy" recommendations, and CSFB initiated coverage with a "strong buy," even though the price had climbed to $90. At $90 per share, Corvis had a market capitalization of $30 billion, despite never having had any revenue. On Friday November 1, 2002, Corvis traded at $0.690 per share at close; there are no stock splits on record.

During the period 1996-2000, 87 percent of analyst initiations at the end of the quiet period were "buys" or "strong buys."

Hypotheses focusing on the allocation of shares

- Government authorities, such as the SEC or the office of the attorney-general of New York, have unearthed cases of conflicts of interest between underwriters and issuers
  - It has been found that if underwriters are given discretion in share allocations, the discretion will not automatically be used in the best interest of the issuer
    - Indeed, there is evidence of investment banks using underpriced share allocations to enrich executives of companies —"friends and family"—with the hope of being awarded investment banking business in return—a practice known as "spinning" and rampant during the period 1996-2000
      - Note than when executives of companies are allocated shares in hot IPOs, the gains from flipping the shares [that is, selling them at the first day of trading] go to the executives whereas the commissions of the investment banking business are paid for by the public shareholder
      - Once an entrepreneur belong to the bank's "friends and family" and has been enriched through deliberate underpricing of other entrepreneurs' IPOs, the cost of having one's own IPO underpriced may deem tolerable the entrepreneur.

One of the investment banks currently under investigation for spinning is Goldman Sachs—innocent until proven guilty

Goldman's hot IPOs

Company	Date of first trade	First day % change	Overall % change	Rating history (price in $) SB=Strong Buy, B=Buy, H=Hold	Latest price
Allegiance Telecom	Jul 1, 1998	−9	−92	SB(66.00)	0.86
Backweb Technologies	Jun 8, 1999	3	−99	B(28.44), H(0.99)	0.19
CoSine Communications	Sep 26, 2000	174	−87	B(47.00)	2.90
eBay	Sep 24, 1998	163	1633*	SB(7.75)	52.01
Engage Technologies	Jul 20, 1999	173	−100	B(20.63), H(2.00)	0.07
eToys	May 20, 1999	282	−100	–	<0.01
Evolving Systems	May 12, 1998	37	−99	B(17.38), SB(8.75)	0.23
Global Crossing	Aug 14, 1998	34	−100	SB(37.56), H(1.45)	0.02
Iasiaworks	Aug 3, 2000	−27	−100	B(11.06)	<0.01
Insweb Corporation	Jul 23, 1999	85	−92	B(147.38)	1.50
iVillage	Mar 19, 1999	241	−98	B(121.50)	0.63
Lucent Technologies	Apr 4, 1996	13	−98	B(26.95), H(4.01)	0.76
PlanetRx.com	Oct 7, 1999	62	−100	B(184.00), H(0.30)	0.08
Portal Software	May 6, 1999	167	−97	SB(62.75), B(9.00), H(8.03)	0.24
Starmedia Network	May 26, 1999	73	−100	B(48.00), H(1.82)	0.01
Tenfold Corporation	May 21, 1999	34	−100	B(49.69), H(7.11)	0.10
theStreet.com	May 11, 1999	215	−89	B(32.63), H(2.75)	2.10
Valueclick	Mar 31, 2000	10	−89	B(11.50)	2.16
Yahoo	Apr 12, 1996	154	786	SB(171.00), B(22.37)	9.46

Sources: House Financial Services Committee; www.ipo.com — * split adjusted

What the House Committee on Financial Services alleges

- Goldman Sachs made shares in hundreds of lucrative initial public offerings (IPOs) available to executives from whom they wanted investment banking business; Credit Suisse First Boston and Salomon Smith Barney, part of Citigroup, also gave clients access to IPOs
- The practice gave these executives an unfair advantage over ordinary investors
- These executives often 'flipped' - sold their shares shortly after the IPO - to lock in profit
- Flipping may have artificially raised the price of some IPOs
- Eight of the 22 IPOs lead-managed by Goldman Sachs that the committee studied rose at least 173 per cent on their first day of trading
- 'There is no equity in the equities markets. I call on every Wall Street firm to show respect for America's individual investors by reforming these corrupt practices immediately.' – *Representative Michael Oxley, Republican from Ohio and committee chairman*
- 'A small circle of preferred clients were given vast access by the investment banks to IPO shares and reaped large profits on the sale of these shares. What is most disturbing is that their profits were gained at the expense of the average investor whose only option was to buy the shares at the oftentimes inflated aftermarket price.' – *Richard Baker, Capital Markets subcommittee chairman*

Source: Financial Times, October 4, 2002.

The immediate post-IPO period

The post-IPO shareholder structure

There evidence that, where bookbuilding is used, institutions receive preferential share allocations—or, in other words, institutions are not rationed as much as retail investors

Institutions are different from retail clients in that they are more likely to be better informed—a characteristic that is of import in the bookbuilding process

There is empirical evidence that clienteles are rather temporary, suggesting that post-IPO corporate control considerations may not be of primary importance in blockbuilding.

In the United States, blockholders—that is, firm outsiders who hold a big chunk of equity—are common prior to the IPO in the form of venture capitalists or leveraged buyout financiers

The venture capitalists typically distribute shares to their limited partners as soon as the post-IPO lockup period ends

Furthermore, the general partners typically also relinquish control via open market sales, rather than selling a strategic block

This supports the aforementioned hypothesis that states that post-IPO corporate control considerations are not be of primary importance in blockbuilding.

The change in shareholder structure from the pre-IPO to the post-IPO period supports the view that IPOs aim at atomistic investors whereas private placements aim at blockholders.

In an empirical study it has been found that when shares are placed more widely than with just a few powerful large shareholders, the entrepreneur is less easily to oust from the company

The same study also finds that management continues to hold on to shares more than other investors, presumably trying to retain control over the company—an indication of private control benefits.

Trading in the secondary market—that is, post-IPO trading

- Once trading commences and the stock price sags, the lead underwriter might attempt to "stabilize" the price by purchasing shares or discouraging selling
  - Price stabilization is the only instance in which the SEC permits active attempts at stock price manipulation.
- Flippers are temporary investors who purchase shares at the IPO and quickly turn it around to sell their shares
  - The new conflict-of-interest theories of underpricing argue that underwriters sometimes allocate shares specifically to investors so that these investors can make a quick profit
  - Furthermore, underwriters desire high aftermarket trading volume because they usually are the prime market makers in the stock
  - On the other hand, when the investment banks fear that the stock might get under pressure in the aftermarket, investment banks discourage flipping through moral suasion—for instance, the threat of withholding future allocations on hot issues—and the imposition of penalty bids
    - In penalty bids, the lead underwriter takes back the selling concession—the concession—from a broker that has allocated shares that are flipped, encouraging the broker to allocate shares on buy-and-hold investors.

The empirical evidence on the effect of the IPO performance in the secondary market—that is, the long-run performance and the extent of short-run underpricing—on the choice of the lead underwriter is mixed at best

Based on interviews of firms that switched underwriters for a follow-on offering after their IPO, it has been shown that the amount of money left on the table—that is, the degree of underpricing—has not been an important factor in the decision to switch

Instead, underwriter prestige—clearly a catch-all term—and the desire to increase analyst coverage for the stock are the two most important determinants of switching.

Underwriters are prohibited from initiating analyst coverage for 25 calendar days after the IPO—the "quiet period"

Usually, the managing underwriters—that is, lead and co-managers—initiate research coverage at the end of the quiet period, usually with a "buy" or "strong buy"

There is empirical evidence that the investment bank's analysts regularly provide "booster shots" in the form of buy recommendations, which are greeted with a positive stock market reaction although these IPOs subsequently underperform

There is empirical evidence that for the period 1996-2000 that where analyst recommendations occur after the quiet period, the market-adjusted return is an average four percent—the market-adjusted return for the remaining IPOs is close to zero

The positive average effect of three percent is difficult to reconcile with market efficiency because the fact that positive recommendations are forthcoming at the end of the quiet period is not a surprise.

For shares not sold in the offering, pre-issue shareholders commit to a specified lockup period during which they agree not to sell shares without the written permission of the lead underwriter

Although there is no statutory minimum, most lockup periods are 180 calendar days in length and almost none is less than 90 days long

There is empirical evidence that the investment bank's analysts regularly provide booster shots in the form of buy recommendations, which are greeted with a positive stock market reaction although these IPOs subsequently underperform

It has been documented that the lead underwriter is typically the dominant market maker for Nadsaq-listed IPOs

The underwriter knows with whom shares are placed and thus has a comparative advantage of contacting investors if there is an order imbalance

There is empirical evidence that market making is a profitable activity for the lead underwriter, with the profits during the first three months of trading amounting to about <u>two percent</u> of the issue size

Also, it has been documented that market-making activity in Nasdaq-listed IPOs continues to be <u>concentrated</u> long after the offering

For smaller IPOs, the lead underwriter's market share declines from almost 100 percent at the inception of trading to less than 50 percent on average about half a year after the offering.

Valuation

A question raised by the difference between the offer price and the first-day market price is whether issuers or the stock market is pricing offerings in line with a company's fundamental

The most common method of valuing IPOs is the use of multiples of comparable, traded companies—the comparable-companies approach

In an empirical study that uses a measure of intrinsic value based on industry-matched sales-to-price and price-to-ebidtda data for a sample of over 2,000 IPOs from the period 1980-1997, it has been shown that IPOs are priced about 50 percent above "comps" (comparable companies)

Not surprisingly, the study finds that the initial overpricing with respect to "comparables" helps predict long-run underperformance.

The Ritter (1991) study on long-run performance of initial public offerings

Reasons why long-run performance of initial public offerings is of interest

Long-run underperformance of IPOs offers opportunities for risk-arbitrage (long-short hedge funds)

Long-run underperformance of IPOs questions the informational efficiency of the IPO market

Evidence of underperformance might lend to support to Robert J. Shiller's (1990, "Speculative Prices and Popular Models," *Journal of Economic Perspectives* 4(2), 55-65) hypothesis that equity markets in general and IPO markets in particular are subject to fads.

If high-volume periods are associated with poor long-run performance, this would indicate that issuers succeed in taking advantage of 'windows of opportunity'.

Data and Methodology

The sample consists of 1,526 initial public offerings in 1975-1984 that meet the following criteria:

The offer price amounts to $1.00 per share or more

The gross proceeds from the IPO equal $1,000,000 or more, measured in 1984 dollars

The offering involved common stock only

The company is listed on the CRSP daily Amex-NYSE or NASDAQ tapes within 6 months of the offer date

The company was taken public by an investment bank.

The sample represents 85.1 percent of the aggregate gross proceeds of all companies that went public in 1975-84

Table 1 shows that the number and proceeds of IPOs are not evenly distributed

Table 1 Distribution of Initial Public Offerings by Year, 1975–84

The number of total offers is based upon *Going Public: The IPO Reporter's* listings, after excluding closed-end mutual funds and real estate investment trusts. Gross proceeds calculations are based upon the amount sold in the United States, including the proceeds from overallotment options, if exercised. No price level adjustments have been made in this table.

	Total of 2,476 Offers		1,526 Offers in Sample		Total Included	
Year	No. of IPOs	Aggregate Gross Proceeds, $ millions	No. of IPOs	Aggregate Gross Proceeds, $ millions	No. of IPOs %	Aggregate Gross Proceeds %
1975	14	264.0	12	262.4	85.7	99.4
1976	33	237.3	28	213.9	84.8	90.1
1977	32	150.6	19	132.3	59.4	87.8
1978	48	247.3	31	218.4	64.6	88.3
1979	78	429.0	53	347.1	67.9	80.9
1980	234	1,408.3	129	1,097.9	55.1	78.0
1981	438	3,200.3	300	2,689.5	68.5	84.0
1982	199	1,335.0	93	1,104.2	46.7	82.7
1983	865	13,247.8	589	12,060.4	68.1	91.0
1984	535	4,237.1	272	2,940.8	50.8	69.4
Total	2,476	24,756.7	1,526	21,066.9	61.6	85.1

Source: Ritter (1993, p. 463).

Only 143 of the 1,526 offers occurred during the first half of the period

Fifty-seven percent of the aggregate gross proceeds in the sample were raised in 1983 alone.

The author uses two measures to gauge the long-run performance of initial public offerings

- First, cumulative average benchmark-adjusted returns (CAR)
  - The adjusted return is the difference between the raw return and the return of a benchmark stock price index; four different benchmark indexes are used.
- Second, three-year buy-and-hold returns (also called 'total returns') for both the IPOs and a set of matching firms
  - The matching firms are represented by American and New York Stock Exchange listed securities that are—to the extent feasible—matched by industry and market capitalization with each IPO.
- Neither performance measure is explicitly adjusted for beta (systematic risk)
  - The authors report that …
    - … the average beta of the IPO companies is greater than unity
    - … the betas of IPO companies, on average, decline over time following the IPO.
  - The betas of the matching firm are also greater than 1.00, but generally somewhat lower than the betas of the IPO firms
    - To the extent that the IPO firms have, on average, somewhat higher betas than the matching firms, the study—by ignoring differences in the market risk premium—might underestimate the degree of long-run underperformance of IPOs: In other words, the results are conservative estimates for IPO underperformance.

Return intervals

There is the initial return period (usually one day), defined as the offering date to the first closing price listed on the CRSP daily return tapes

There is the aftermarket period, defined as the three years after IPO exclusive of the initial return period.

The initial return period is defined to be month 0, and the aftermarket period includes the following 36 months where months are defined as successive 21-trading-day periods relative to the IPO date

For IPOs that are delisted before their 3-year anniversary, the aftermarket period is truncated.

For the adjusted returns, the author uses four different benchmarks

CRSP value-weighted NASDAQ index

CRSP value-weighted Amex-NYSE index

Listed firms matched by industry and size

Index of the smallest size decile of the New York Stock Exchange.

Calculation of benchmark-adjusted returns

The benchmark-adjusted return for stock i in month t, $r_{i,t}$ is defined as

$$ar_{i,t} = r_{i,t} - r_{m,t}$$

where $r_{m,t}$ is the return of the benchmark portfolio, m (market).

The average benchmark-adjusted return on a portfolio of n stocks for event month t is the (equally-weighted) arithmetic mean of the n benchmark-adjusted returns:

$$\mathrm{AR}_t = \frac{1}{n} \cdot \sum_{i=1}^{n} ar_{i,t}$$

The cumulative benchmark-adjusted return, CAR, from event month q to event month s is the sum of the average benchmark-adjusted returns:

$$\text{CAR}_{q,s} = \sum_{t=q}^{s} \text{AR}_t$$

When a company in portfolio p is delisted from the CRSP data, the portfolio return for the next month is an (equally-weighted) average of the remaining companies in the portfolio, which implies monthly rebalancing.

Calculation of three-year holding period returns

The three-year holding period return is the return from a buy-and-hold strategy where a stock is purchased at the first closing market price after going public and held until the earlier of its three-year anniversary or its delisting in month $k+1$:

$$R_i = \prod_{t=1}^{t=\min\{36;k\}} (1 + r_{i,t})$$

where $r_{i,t}$ is the raw return on stock i in event month t

For ease of interpretation, the author calculates wealth relatives as a performance measure, defined as

$$WR = \frac{1 + \text{average 3-year total return on IPOs}}{1 + \text{average 3-year total return on matching firms}}$$

A wealth relative of greater than unity is indicative of IPOs outperforming a portfolio of matching companies; a wealth relative of less than unity indicates that IPOs underperformed.

Empirical findings

Abnormal returns for initial public offerings in 1975-1984

Table 2 reports the average matching firm-adjusted returns, AR_t and cumulative average matching firm-adjusted returns, $CAR_{1,t}$, for each of the 36 months after the offering date, excluding the initial return (which, usually, is the return on the first day of trading)

Thirty-one of the 36 monthly average adjusted returns are negative; about a third of them is statistically significant

After a slight increase in the first two months of seasoning, the cumulative average adjusted return declines steadily to -29.13 percent (*t*-statistic: -5.89) by the end of month 36

Table 2 Abnormal Returns for Initial Public Offerings in 1975–84

Average matching firm-adjusted returns (AR_t) and cumulative average returns ($CAR_{1,t}$), in percent, with associated t-statistics for the 36 months after going public, excluding the initial return. The number of firms trading begins at less than 1,526 because some firms have a delay of more than one month after going public before being listed. $AR_t = 1/n_t \Sigma_{i=1}^{n_t}(r_{ipo,it} - r_{match,it})$ where $r_{ipo,it}$ is the total return on initial public offering firm i in event month t, and $r_{match,it}$ is the total return on the corresponding matching firm. The t-statistic for the average adjusted return is computed for each month as $AR_t \cdot \sqrt{n_t}/sd_t$, where AR_t is the average matching firm-adjusted return for month t, n_t is the number of observations in month t, and sd_t is the cross-sectional standard deviation of the adjusted returns for month t. The cross-sectional standard deviations vary from a low of 19.02 percent in month 10 to a high of 25.24 percent in month 16. The t-statistic for the cumulative average adjusted return in month t, $CAR_{1,t}$, is computed as $CAR_{1,t} \cdot \sqrt{n_t}/csd_t$, where n_t is the number of firms trading in each month, and csd_t is computed as $csd_t = [t \cdot var + 2 \cdot (t - 1) \cdot cov]^{1/2}$, where t is the event month, var is the average (over 36 months) cross-sectional variance, and cov is the first-order autocovariance of the AR_t series. Var has a value of 0.04453 (21.10 percent squared) and cov has a value of 0.02097, representing an autocorrelation coefficient of 0.471.

Month of Seasoning	Number of Firms Trading	AR_t %	t-stat	$CAR_{1,t}$ %	t-stat
1	1,512	0.38	0.63	0.38	0.70
2	1,514	1.49	2.81	1.88	2.02
3	1,517	−0.12	−0.24	1.75	1.46
4	1,518	−1.07	−2.21	0.69	0.48
5	1,519	−0.81	−1.63	−0.12	−0.08
6	1,519	−0.55	−1.06	−0.67	−0.38
7	1,518	−1.59	−3.13	−2.27	−1.18
8	1,516	−1.10	−2.21	−3.37	−1.63
9	1,514	−1.73	−3.38	−5.10	−2.31
10	1,513	−1.63	−3.32	−6.72	−2.88
11	1,508	−1.59	−3.08	−8.32	−3.39
12	1,501	−1.91	−3.66	−10.23	−3.97
13	1,496	−0.32	−0.56	−10.55	−3.92
14	1,492	−0.82	−1.60	−11.37	−4.06
15	1,486	−1.19	−2.30	−12.56	−4.32
16	1,478	−1.26	−1.92	−13.82	−4.59
17	1,469	−0.47	−0.85	−14.29	−4.58
18	1,463	−0.49	−0.88	−14.78	−4.59
19	1,449	0.37	0.61	−14.42	−4.43
20	1,440	0.30	0.55	−14.11	−4.12

Table 2 (*Continued*)

Month of Seasoning	Number of Firms Trading	AR_t %	t-stat	$CAR_{1,t}$ %	t-stat
21	1,429	−0.94	−1.66	−15.05	−4.27
22	1,416	−0.20	−0.33	−15.25	−4.21
23	1,403	−0.56	−0.92	−15.80	−4.24
24	1,397	−1.09	−1.97	−16.89	−4.43
25	1,388	0.30	0.50	−16.59	−4.25
26	1,372	−0.26	−0.44	−16.85	−4.20
27	1,354	−1.66	−2.87	−18.51	−4.50
28	1,347	−1.02	−1.72	−19.54	−4.65
29	1,339	−0.97	−1.84	−20.51	−4.78
30	1,324	−1.51	−2.74	−22.01	−5.01
31	1,309	−1.02	−1.57	−23.03	−5.13
32	1,296	−0.63	−1.00	−23.66	−5.16
33	1,283	−1.31	−2.16	−24.96	−5.33
34	1,270	−1.39	−2.39	−26.35	−5.52
35	1,260	−1.10	−1.89	−27.45	−5.64
36	1,254	−1.67	−2.80	−29.13	−5.89

Source: Ritter (1993, p. 467-68).

Figure 1 plots the matching firm-adjusted CAR shown in Table 1 (long dashes), the initial return included

There is an abnormal return of 14.32 percent on the first trading day

After about 18 months, the cumulative return turns negative

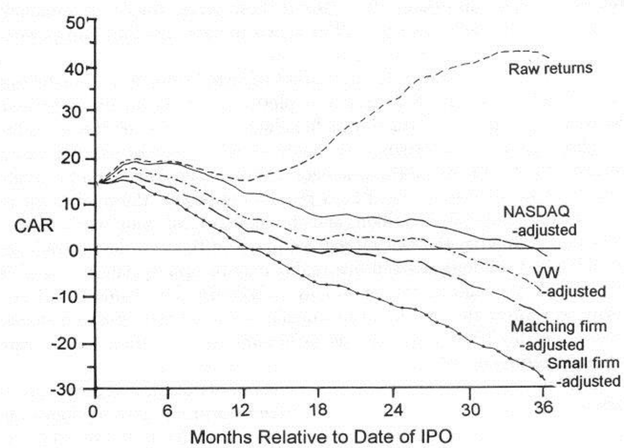
After 36 months, the IPO has underpeformed the matched-firms portfolio by roughly 15 percentage points.

Figure 1 also plots the CAR adjusted by benchmarks other than the matched sample

Because many of the IPOs are small caps, a small cap benchmark portfolio might be appropriate

Using an equally-weighted index of small stock returns, as represented by the lowest decile of market capitalization stocks trading on the NYSE, the months 1-36 cumulative average small-cap adjusted return equals -42.21*%*

Figure 1 Cumulative average adjusted returns for an equally-weighted portfolio of 1,526 initial public offerings in 1975–84, with monthly rebalancing.



Five CAR series are plotted for the 36 months after the IPO date: 1) no adjustment (raw returns), 2) CRSP value-weighted NASDAQ index adjustment (NASDAQ-adjusted), 3) CRSP value-weighted Amex-NYSE index adjustment (VW-adjusted), 4) matching firm adjustment (matching firm-adjusted), and 5) lowest decile of NYSE market capitalization index adjustment (small firm-adjusted). Month 0 is the initial return interval.

Source: Ritter (1993, p. 469).

Table 3 reports the distribution of three-year buy-and-hold period returns for both the 1,526 IPOs and the matching firms

- The median IPO three-year return equals -16.67%, in contrast with 38.54% for the median matching firms
- The mean IPO three-year return equals 34.47%, compared to a mean of 61.86% for the matching firms
  - The distribution of holding period returns for IPOs is more skewed than those for the matching firms

Table 3 Distribution of Three-Year Holding Period Returns, Exclusive of the Initial Return, for 1,526 Initial Public Offerings and Matching Firms in 1975–84

Three-year holding period returns are calculated as $[\Pi_{t=1}^{756}(1 + r_{idt}) - 1] \times 100\%$ where r_{idt} is the daily return on stock *i*, with the CRSP daily NASDAQ returns tape and the daily Amex-NYSE returns tape being the source of the daily returns. For initial public offerings that were delisted before the 3-year anniversary, the total return is calculated until the delisting date. If the initial return period lasted for more than 1 day, the total return is calculated from the first CRSP-reported closing price until the 756th trading day after the IPO. The corresponding matching firm's total return is calculated over the same truncated return interval. If the matching firm is delisted early, a second (and possibly third) matching firm's return is spliced onto the first matching firm. For firms with no dividends and no stock splits the total return corresponds to $[(P_3/P_t) - 1] \times 100\%$ where P_3 is the price on the 3-year anniversary, and P_t is the first closing market price after the IPO.

	Three-Year Holding Period Total Return, in Percent	
Rank	Initial Public Offerings	Matching Firms
1 (lowest)	−99.02	−94.59
77	−92.27	−61.11
153	−85.80	−44.53
229	−79.16	−31.58
306	−72.98	−19.05
382 (25th percentile)	−66.25	−9.49
458	−57.98	0.00
535	−48.06	8.79
611	−38.28	18.75
687	−28.20	27.67
764 (median)	−16.67	38.54
840	−2.54	51.14
916	13.25	63.14
992	29.07	75.82
1,069	46.85	87.56
1,145 (75th percentile)	69.59	103.16
1,222	99.96	120.42
1,298	138.03	148.86
1,374	205.33	187.01
1,450	320.53	240.78
1,526 (highest)	3,964.43	1,268.56
Mean	34.47	61.86

Source: Ritter (1993, p. 472).

Table 4 displays a breakdown of initial returns and three-year buy-and-hold returns by gross proceeds

- All gross proceeds categories display long-run underperformance
- Small offers have the highest average matching-firm adjusted initial returns and the worst aftermarket performance
  - This result is reminiscent of the finding of N. Chopra, J. Lakonishok and J.R. Ritter (1992; "Measuring Abnormal Performance: Do Stocks Overreact?" *Journal of Financial Economics* 31, 235-268) that states that overreaction is particularly strong among small firms, which are predominantly owned by individuals
  - Note that these small firms do not fully exploit the market's over-optimism, as to avoid lawsuits.

Table 4 also reports <u>median</u> initial and aftermarket returns

For the initial returns, the median is a positive 4.61 percent, with only 368 of the 1,526 offers (22.2 percent) having negative adjusted initial returns

Table 4 Mean Performance Measures for 1,526 IPOs in 1975–84 Categorized by Gross Proceeds

Gross proceeds are measured in dollars of 1984 purchasing power using the U.S. GNP deflator. Initial returns are computed as $r_{ipo} - r_{matching\ firm}$ over the initial return interval (one day for 1,203 of the 1,526 firms). The three-year holding period return is calculated excluding the initial return. For IPOs that are delisted prior to their three-year anniversary the matching firms' return is ended on the same date as the IPO. Total returns include both capital gains and dividends. The *wealth relative* is the ratio of one plus the mean IPO 3-year holding period return (not in percent) divided by one plus the mean matching firm 3-year holding period return (not in percent). For the smallest gross proceeds category, 1.1794/1.6754 = 0.704.

		Excluding Initial Returns				
		Average 3-Year Holding Period Total Return			Sample Size	
Gross Proceeds, $	Average Adjusted Initial Return %	IPOs %	Matching Firms %	*Wealth Relative*	Month 0	Month 36
1,000,000–2,999,999	27.45	17.94	67.54	0.704	221	146
3,000,000–4,999,999	18.00	20.89	58.72	0.762	296	238
5,000,000–9,999,999	11.28	40.06	69.87	0.825	379	316
10,000,000–14,999,999	7.51	46.25	55.99	0.938	211	183
15,000,000–24,999,999	10.09	43.97	50.56	0.956	200	179
25,000,000–353,950,260	9.96	39.81	62.50	0.860	219	193
All (mean)	14.06	34.47	61.86	0.831	1,526	1,254
All (median)	4.61	−16.67	38.54	0.601	1,526	1,254

Source: Ritter (1993, p. 473).

Table 6 shows that IPOs—in a given period—tend to be concentrated in certain industries

In the period 1975-84, there is concentration of IPOs in the financial services and the airline industries due to deregulation

Most financial institutions that went public were mutual savings banks and mutual savings and loan associations that converted to stock corporations in 1983-84 in response to a 1982 regulatory change; their median age at the time of the IPO is 49 years.

There is also a concentration in the oil and gas industry and in the computer and communications industries

The median age of companies that went public in the oil and gas industry and the computer and communications industries varies between two years (oil and gas) and seven years (communications and electronic equipment)

Table 6 Mean and Median Sales, Gross Proceeds, and Age of 1,526 Sample Offers Categorized by Industry

Both sales and gross proceeds are expressed in terms of dollars of 1984 purchasing power. Sales are measured as 12-month revenues for the most recent 12-month period prior to going public. Gross proceeds are measured including, for firm commitment offerings, the proceeds from overallotment options, if exercised. The age of the issuing firm is measured as the calendar year of going public minus the calendar year of founding. The year of founding is the same or earlier than the year of incorporation or reincorporation. The 39 firms with a founding date prior to 1901 have their age computed as the offer year minus 1901.

Industry	SIC Codes	Number of Offers	Annual Sales, $ Millions		Gross Proceeds, $ Millions		Age of Issuing Firm	
			Mean	Median	Mean	Median	Mean	Median
Computer manufacturing	357	144	18.78	14.36	21.40	13.46	6.20	5
Communications and electronic equipment	366, 367	138	14.16	8.26	11.25	6.21	9.28	7
Oil and gas	131, 138 291, 679	127	19.05	0.53	9.57	6.33	4.83	2
Financial institutions (banks and S&L's)	602, 603 612, 671	125	120.20	49.43	27.41	12.00	43.13	49
Computer and data processing services	737	113	16.40	11.50	13.98	9.83	9.55	7
Optical, medical, and scientific instruments	381–384	111	10.89	2.23	9.29	4.66	8.08	5
Retailers	520–573 591–599	70	74.70	34.74	17.28	11.10	12.89	7
Wholesalers	501–519	63	56.58	13.41	12.32	4.72	9.65	6
Restaurant chains	581	54	34.54	10.98	10.34	5.55	7.33	4
Health care and HMOs	805–809	50	35.20	7.42	14.96	6.80	4.88	3
Drugs and genetic engineering	283	44	21.14	1.98	19.70	11.55	7.68	3
Miscellaneous business services	739	42	14.27	2.38	7.75	4.93	8.55	8
Airlines	451	25	20.65	14.33	11.68	6.00	7.84	4
All other firms	—	420	61.24	18.07	15.04	7.32	14.58	8
All firms	—	1,526	42.82	11.55	15.06	7.59	12.66	6

Source: Ritter (1993, p. 476-77).

Table 7 breaks down aftermarket performance by industry

- Matching-firm adjusted long-run performance of IPOs (the wealth relative) varies widely across industries and is present in all but three of the fourteen industries
  - Financial institutions, which have a median age of 49 years at the time of going public (Table 6), have the highest wealth relative; the wealth relative is greater than one
  - IPOs in the oil and gas industry, which have a median age of only two years at the time of going public (Table 6), substantially underperformed their peers, exhibiting the lowest wealth relative.

Table 7 Mean Performance Categorized by Industry

The *wealth relative* is the ratio of one plus the mean IPO 3-year holding period return (not in percent) divided by one plus the mean matching firm 3-year holding period return (not in percent).

		Excluding Initial Returns		
		Average 3-Year Holding Period Total Return		
Industry	Average Matching Firm-Adjusted Initial Return %	IPOs %	Matching Firms %	Wealth Relative
Computers	13.67	19.22	47.84	0.806
Electronic equipment	14.59	29.93	61.46	0.805
Oil and gas	30.92	−43.86	34.67	0.417
Financial institutions	3.69	128.21	59.23	1.433
Computer services	16.07	13.13	50.38	0.752
Scientific instruments	20.96	18.14	72.20	0.686
Retailers	7.60	54.05	113.63	0.721
Wholesalers	16.95	1.42	47.14	0.689
Restaurant chains	13.51	73.86	82.36	0.953
Health care	14.12	36.93	53.25	0.894
Drugs	14.63	121.69	91.96	1.155
Miscellaneous services	10.20	26.61	80.50	0.701
Airlines	6.26	61.62	42.93	1.131
All other firms	11.13	33.40	64.24	0.812
All firms	14.06	34.47	61.86	0.831

Source: Ritter (1993, p. 478).

In Table 8, IPOs are categorized by their year of issuance

Table 8 Performance Categorized by Year of Issuance for Initial Public Offerings in 1975–84

The average real gross proceeds, measured in dollars of 1984 purchasing power, is computed as the product of the U.S. GNP Deflator index and the average nominal gross proceeds. The *wealth relative* is the ratio of one plus the mean IPO 3-year holding period return (not in percent) divided by one plus the mean matching firm 3-year holding period return (not in percent), exclusive of the initial return.

		Average Gross Proceeds, $ Millions				Excluding Initial Returns: Average 3-Year Holding Period Total Return		
Year	GNP Deflator	Nominal	Real	Number of Issues	Average Matching Firm-Adjusted Initial Return %	IPOs %	Matching Firms %	Wealth Relative
1975	1.76	21.87	38.49	12	−5.24	59.44	52.51	1.045
1976	1.67	7.64	12.76	28	6.38	122.58	124.11	0.993
1977	1.58	6.96	11.00	19	8.21	188.35	54.72	1.864
1978	1.47	7.05	10.36	31	31.78	134.60	97.37	1.189
1979	1.38	6.55	9.04	63	22.06	75.98	71.76	1.025
1980	1.26	8.51	10.72	129	38.27	46.28	68.56	0.868
1981	1.15	8.96	10.30	300	9.98	5.26	60.85	0.654
1982	1.08	11.87	12.82	93	15.18	26.07	119.92	0.573
1983	1.04	20.48	21.30	589	12.56	21.31	52.88	0.793
1984	1.00	10.81	10.81	272	8.40	52.03[a]	47.91	1.028[a]
All	—	13.81	15.06	1,526	14.06	34.47	61.86	0.831

[a] If one outlier (TCBY, Inc.) is removed, the average 3-year raw return falls to 37.59% and the 3-year *wealth relative* falls to 0.930.

Source: Ritter (1993, p. 481).

The results show that the long-run underperformance is not as general a phenomenon as the short-run underpricing that has been widely documented

The wealth relatives are less than one for only five of the ten sample years

Because the volume of new issues was much heavier in the early 1980s than in the late 1970s, with all issues weighted equally, the mean wealth relative is a mere 0.831.

Also, there is a negative relation between annual volume and subsequent aftermarket performance

- This finding lends support to the hypothesis that companies choose to go public when investors are willing to pay high 'multiples' (price-earnings or market-to-book ratios), reflecting optimistic expectations about earnings growth

    - The poor subsequent aftermarket performance then is the result of a rough awakening to disappointing cash flows.

Table 9 offers a breakdown of aftermarket performance by age of the company at the time of going public

- The age of the firm is calculated as the year of the offer minus the year of founding
- There is a monotone, negative relation between age and initial return
  - This finding is consistent with the notion that risky issues require higher initial returns and that (high) age is a proxy for (low) risk.
- There is a monotone, negative relation between age and aftermarket performance (wealth relative)
  - This finding is consistent with the notion that investors are overoptimistic about the earnings prospects of young 'growth firms'.
- In Panel B, Table 9, the author excludes oil and gas firms and financial institutions
  - Remember that the oil and gas firms are, simultaneously, in the extreme tails of initial returns (Table 7), aftermarket performance (ditto) and age (Table 6).
  - Panel B shows that the results from Panel A hold up even when the financial services and oil and gas industries are left out.

Table 9 Aftermarket Performance Categorized by the Age of the Issuing Firm

Panel A includes all 1,526 IPOs. Panel B includes the 1,274 IPOs remaining after excluding the two industries with the most extreme wealth relatives: oil and gas (primarily very young firms) which did poorly, and financial institutions (primarily very old firms) which did well. Oil and gas firms are defined as firms with SIC codes of 131, 138, 291, and 679, representing oil and gas exploration, production, servicing, refining, and holding companies. Financial institutions are defined as firms with SIC codes of 602, 603, 612, and 671, representing commercial banks, savings banks, savings and loans, and bank holding companies. The *wealth relative* is the ratio of one plus the mean IPO 3-year holding period return (not in percent) divided by one plus the mean matching firm 3-year holding period return (not in percent).

			Excluding Initial Returns		
			Average 3-Year Holding Period Total Return		
Age in Years	Sample Size	Average Matching Firm-Adjusted Initial Return %	IPOs %	Matching Firms %	Wealth Relative
Panel A: All 1,526 firms					
0–1	252	29.42	5.34	68.98	0.623
2–4	381	14.51	15.69	48.69	0.778
5–9	328	13.15	28.47	62.33	0.791
10–19	312	9.05	40.74	66.70	0.844
20–up	253	5.42	91.81	68.03	1.142
Panel B: Excluding oil and gas firms and financial institutions					
0–1	177	23.87	16.19	76.31	0.659
2–4	338	14.87	19.22	53.20	0.778
5–9	305	13.71	33.01	65.47	0.804
10–19	300	9.32	42.97	67.66	0.853
20–up	154	5.41	63.76	70.37	0.961

Source: Ritter (1993, p. 481).

Table 10 shows the results of a regression analysis

- In multivariate regressions, it is possible to disentangle multiple, simultaneous influences
- The author regresses the (unadjusted) three-year buy-and-hold return on …
  - … the initial return (IR)
  - … the log of 'age plus one year' (Log(1+age))
  - … the return on the CRSP value-weighted Amex-NYSE market index (Market)
  - … the volume of IPOs in the calendar year of issuance (Vol)
  - … 0/1 ('dummy') variables for the oil and gas and financial services industries (Oil and Bank, respectively).
- The regression coefficients have the expected signs
  - With the exception of the initial return, all the coefficients are statistically significant at the conventional levels
  - The author might have wanted to leave out the initial return variable (IR) because the initial return might be explained by the same variables that explain the dependent variable of this regression (which is the long-run performance)
    - The insignificance of IR might be due to misspecification (multicollinearity).
- The regression coefficient for the IPO volume of -0.109, for instance, indicates that the difference in 3-year total returns for a firm going public in a low volume year as 1976 (33 offerings) rather than in a high volume year such as 1983 (865 offerings) is 90.7 percentage points ($(865-33)\cdot 0.109 \approx 90.7$)).

Table 10 Ordinary Least Squares Regression Results with the Three-year Total Return as the Dependent Variable, for 1,526 IPOs in 1975–84

$\text{Return}_i = b_0 + b_1\text{IR}_i + b_2\text{Log}(1 + \text{age}_i) + b_3\text{Market}_i + b_4\text{Vol}_i + b_5\text{Oil}_i + b_6\text{Bank}_i + e_i$. Return_i is the raw three-year return, measured from the first aftermarket closing price to the earlier of the three-year anniversary or its CRSP delisting date. IR_i is the market-adjusted initial return, calculated using the CRSP value-weighted index of Amex-NYSE stocks as the market index. $\text{Log}(1 + \text{age}_i)$ is the natural logarithm of one plus the difference between the year of going public and the year of founding, with firms founded before 1901 assumed to be founded in 1901. Market_i is the CRSP value-weighted market return for the same return interval as the dependent variable. Vol_i is the annual volume of IPOs in the year of issuance, divided by 100. The gross number of IPOs, given in Table 1, is used. Oil_i is a 0, 1 dummy variable taking on the value of 1 if the issuing firm has an SIC code of 131, 138, 291, or 679, representing oil and gas production, exploration, refining, and service companies, or oil and gas holding companies. Bank_i is a 0, 1 dummy variable taking on the value of 1 if the issuing firm has an SIC code of 602, 603, 612, or 671, representing banks, savings and loans, and associated holding companies. Bootstrapped p-values are in parentheses.

Panel A: Parameter estimates

Intercept	IR	Log(1 + age)	Market	Vol	Oil	Bank	R^2_{adjusted}
0.238	−0.206	0.127	0.841	−0.109	−0.765	0.825	0.070
(0.186)	(0.143)	(0.010)	(0.001)	(0.001)	(0.001)	(0.001)	

Panel B: Summary statistics of variables

Variable	Mean	Median	Standard Deviation
Return	0.345	−0.167	1.902
IR	0.141	0.040	0.309
Log(1 + age)	2.009	1.946	1.079
Market	0.566	0.580	0.246
Vol	5.520	5.350	2.831
Oil	0.083	0.000	0.276
Bank	0.082	0.000	0.274

Source: Ritter (1993, p. 483).

Summary

- The average holding period return of a sample of 1,526 IPOs of common stock in 1975-1984 is 34.47 percent in the three years after going public, where the holding period return is measured from the closing price on the first day of trading to the market price on the 3-year anniversary
  - A control sample of 1,526 listed stocks, matched by industry and market value, generates an average total return of 61.86 percent over this same 3-year holding period
    - A strategy of investing in IPOs at the end of the first day of public trading and holding them for three years would have left the investor with only 83 cents relative to each dollar from investing in a group of matching firms listed on the American and New York Stock Exchanges
      - Young companies and companies going public in heavy-volume years showed particularly poor long-run performance.
- Combined with evidence on high initial returns of IPOs (on the first day of trading), the evidence on long-run underperformance indicates that the offering price is not too low; rather, the closing price on the first day of trading is too high
  - Fads—i.e., periods of over-optimism—are a possible explanation of the observed short-run and long-run returns patterns
    - The concentration in volume in certain years and industries indicates that companies take advantage of 'windows of opportunity' by going public near the peak of industry-specific fads.

Criticism

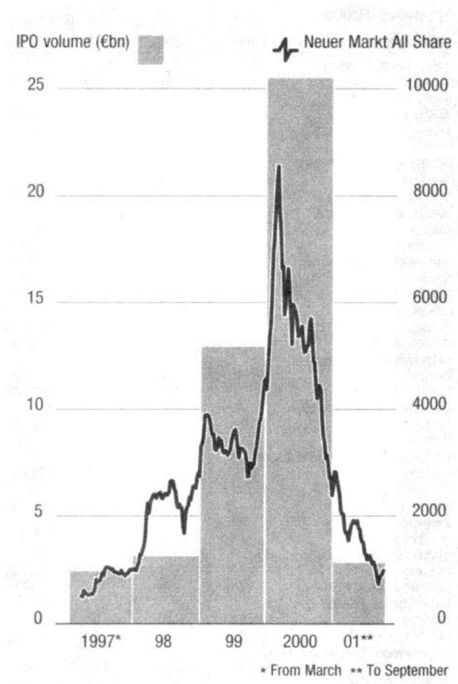
- Methodological issues have been raised about how to properly measure benchmark-adjusted returns when the returns of IPOs overlap in calendar time—as is the case in the approach chosen in the table below
  - Overlap in calendar time is associated with positive cross-sectional dependence because of unpriced industry factors in returns, which is to say that the observations are not statistically independent
  - A way out would be to estimate an asset pricing model in calendar time, such as the Fama-French three-factor model—the factors are the CAPM beta, market capitalization, and market-to-book value—augmented by a fourth factor—an IPO indicator variable
    - The problem inherent in the Fama-French model is the assumption of efficient markets or, in other words, any factor that generates abnormal returns is—by assumption—a risk factor
      - Then again, if there is indeed a behavioral component in the long-run performance of IPOs, the Fama-French model would not discover it.

Sources of long-run underperformance

- Short-selling constraints, along with heterogeneous beliefs among investors about the value of the company might be one reason why IPOs underperform over the long haul
  - The most optimistic buyers buy the IPO
    - Over time, as the variance of opinion decreases, the marginal investor's valuation will converge toward the mean valuation, and the share price will shall
  - This hypothesis is consistent with the drop in share price at the end of the lockup period—a phenomenon documented in several empirical studies.
- Another explanation is that successful IPOs lead to more IPOs
  - Thus, the last large group of IPOs would underperform and be a relatively large fraction of the sample
  - Clearly, this hypothesis does not suffice in explaining long-run underperformance because underperformance is not restricted to just the last cohort—or couple of last cohorts—of a given host-issue market.

Casual evidence

Stock market (over-)valuation and IPO volume



Source: Financial Times, October 29, 2001.

Long-term IPO performance of the U.S. top 10 performers ranked by initial returns (first-day gains), 1998 through early 2000

TOP 10 IPO DEBUTS

Date	Issuer	First Day Gain	Current Drop
12/99	VA Linux Systems	697.5%	-97%
11/88	Theglobe.com	605.56%	-96%
9/99	Foundry Networks	525%	-85%
2/00	WebMethods	507.5%	-90%
12/99	FreeMarkets	483.3%	-86%
11/99	Cobalt Networks	482.4%	---*
1/99	MarketWatch.com	473.5%	-80%
10/99	Akamai Technologies	458.4%	-94%
11/99	CacheFlow Inc.	426.6%	-98%
3/00	Crayfish Co. Ltd.	422.7%	-90%

*Sun Microsystems bought Cobalt for $2B last year
SOURCE: DEALOGIC

Source: http://money.cnn.com, posted August 25, 2001.

[Blank page]

Empirical Studies

14. Stock Market Underreaction (Drift)

Reference:

Bernard, Victor L., and Jacob K. Thomas (1989) "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?" *Journal of Accounting Research* 27, Supplement, 1-36.

Note: All figures and tables in this chapter are from Bernard and Thomas (1989)

Ball and Brown (1968, "An Empirical Evaluation of Accounting Income Numbers," *Journal of Accounting Research*, Autumn, 159-78) were the first to note that even after earnings are announced, estimated cumulative abnormal returns continue to drift up for "good news" companies and down for "bad news" companies

Note that U.S. corporations announce earnings <u>quarterly</u> (in the first week of the next quarter), but also issue <u>earnings warnings</u>, typically in the last weeks of the current quarter

Remember that <u>abnormal returns</u> are returns above and beyond the excess (i.e., risk-adjusted) return

Unfortunately, in many studies abnormal returns are measured as returns in excess of the market return, which holds the implicit assumption that the stock (or portfolio) in question has a <u>unit CAPM beta</u>

If abnormal return is measured as excess return, abnormal return is underestimated for stocks (or portfolios) with CAPM betas less than unity, and overestimated for CAPM betas greater than unity.

The findings by Ball and Brown (1968) have been confirmed in many studies, among them Foster, Olsen, Shevlin (1984, "Earnings Releases, Anomalies, and the Behavior of Security Returns," *Accounting Review*, October, 574-603)

- Foster, Olsen, and Shevlin (1984) estimate that over the 60 trading days subsequent to earnings announcements, a long position in stocks with unexpected earnings in the highest decile, combined with a short position in stocks in the lowest decile, yields an annualized return of about 25 percent before transactions costs
  - Competing explanations for <u>post-earnings announcement drift</u> fall into two categories
    - First, at least a portion of the price response to new information is delayed
      - Investors might not update beliefs fully to new information
      - Transactions costs might prevent investors from incorporating new information into securities prices immediately.

Second, the capital-asset pricing model (CAPM), which serves as a benchmark for determining abnormal returns, might be misspecified in that certain risk factors are not accounted for

As a result, the so-called abnormal return is nothing more than fair compensation for bearing risk that is priced in the market but not captured by the CAPM

(Remember the multi-factor models from the chapter "Predictability of Stock Returns," and the findings of Eugene F. Fama and Kenneth R. French (1992, "The cross-section of expected stock returns," *Journal of Finance* 47, 427-66) that firm size and price-to-book ratio explain cross-sectional differences in expected excess returns.)

In the case of post-earnings-announcement drift, this explanation requires that companies with higher than expected earnings become more risky or companies with lower than expected earnings become less risky (or both)—on some unrecognized dimension.

The study by Bernard and Thomas (1989), henceforth "the authors"

- The study suggests that post-earnings-announcement drift …
  - … is difficult to reconcile with plausible explanations of changes in firm risk
  - … is consistent with a delayed response of prices to information
    - The delayed response can only partially be explained by transaction costs, which suggests that it is due at least in part to incomplete updating of beliefs about future earnings.
- Sample selection
  - The authors' sample includes 84,792 firm-quarters of data for NYSE (New York Stock Exchange) / AMEX (American Stock Exchange) stocks for 1974-86
  - To be included in the sample, stocks have to be listed on the CRSP files and the company's earnings before extraordinary items and discontinued operations have to be available for at least ten consecutive quarters on Compustat
    - The NYSE/AMEX sample includes only companies that appeared on any of the Compustat files released from 1982 through 1987
      - Since companies included in earlier files, but dropped from Compustat before 1982, are excluded from the sample, there is potential for a survivorship bias in the sample years 1974-81
      - One the other hand, the authors point out that their conclusions are insensitive to whether they include or exclude "nonsurvivors" dropped from the Compustat files between 1982 and 1987
        - Note that this argument of the authors' has some currency because nonsurvivors are likely to be companies with negative earnings surprises and consequently included in the short (rather than the long) position of the authors' portfolio.

Estimation of abnormal returns

The abnormal return of company j on day t, AR_{jt}, is defined as

$$AR_{jt} = R_{jt} - R_{pt}$$

where R_{jt} is the raw return for company j on trading day t

R_{pt} is the equally weighted mean return for trading day t on the NYSE/AMEX firm size decile that company j belongs to at the beginning of the calendar year in question; firm size is measured by the market value of common stock.

Observations are excluded from the analysis if the return for the earnings announcement day was missing on CRSP or if the CRSP returns series do not encompass the 160 trading days surrounding the earnings announcement

Note that the abnormal return as defined by the authors is in fact nothing other than an excess return

The authors' definition of abnormal return does not control for differences in systematic risk across stocks, as captured by the CAPM beta

In other words, if companies with positive earnings surprises exhibit systematically different systematic risk than companies with negative earnings surprises, the calculated returns on self-financing long-short portfolios might be biased.

The authors use size deciles to eliminate the small firm effect on expected return (see the chapter "Predictability of Stock Returns")

Estimation of standardized unexpected earnings (SUE)

To identify earnings surprises, the authors follow a procedure suggested by Foster, Olsen, and Shevlin (1984) who measure surprises relative to earnings forecasts derived from historical data

SUE's are then calculated by normalizing the difference between forecast and actual earnings by the standard deviation of forecast errors over the estimation period

The earnings forecasting model assumes that earnings follow a first-order autoregressive process in seasonal differences

In other words, the authors assume that there is positive first-order serial correlation in changes in earnings, where changes in earnings are measured as differences to the same quarter a year ago.

Returns on self-financing long-short portfolios

To gauge market efficiency, the authors calculate cumulative abnormal returns (CAR's) on two kinds of self-financing long-short portfolios over windows of 60 trading days

FOS control portfolio SUE strategy

Following Foster, Olsen, and Shevlin [FOS] (1984), the authors calculate CAR's for a portfolio that goes long on the highest SUE decile and short on the lowest SUE decile

The returns on the long (or short) positions are differences to size-control portfolios, as defined above

The authors calculate CAR's for these positions by adding up daily returns (rather than compounding them)

Adding up returns implicitly assumes daily portfolio rebalancing.

Continuously balanced SUE strategy

Given that daily portfolio rebalancing can hardly be a profitable investment strategy, the authors also provide buy-and-hold (i.e., compounded) CAR's on a portfolio that goes long on companies with positive unexpected earnings announcements and matches each of these "good news" companies individually—dollar for dollar—with (one or more) "bad news companies" of the same size category (of which there are three)

For cases in which no "bad news" companies become available on a day in which a "good news" company needs a match, the authors start the 60-trading-day clock on the day the next "bad news" observation (from the same size category) becomes available.

Empirical results

The empirical results provided by the authors confirm the findings of Foster, Olsen, and Shevlin (1984) who show that (i) there is post-earnings-announcement drift in stock prices, (ii) the magnitude of post-earnings-announcement drift varies negatively with firm size, and (iii) there is persistence (longevity) in post-earnings-announcement drift

The figure below (authors' Figure 2) presents CAR plots, where companies are assigned to (ten) portfolios on the basis of SUE's in the 60-trading-day window <u>prior to the announcement</u>

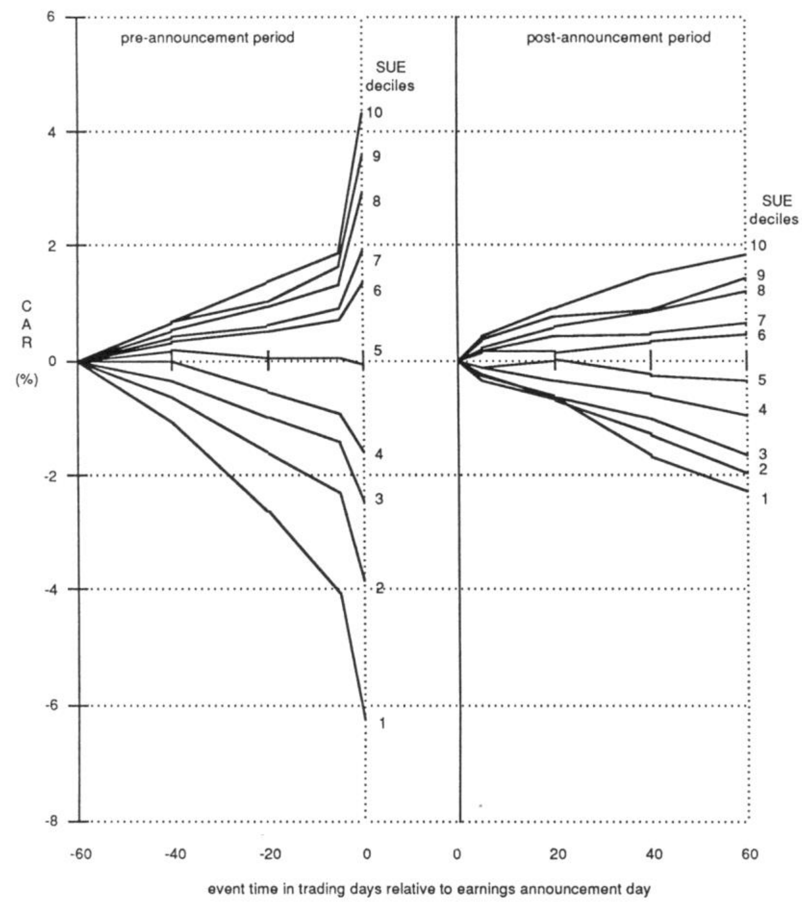


FIG. 2.—Cumulative abnormal returns (*CARs*) for *SUE* portfolios: all announcements. Earnings announcements are assigned to deciles based on standing of standardized unexpected earnings (*SUE*) relative to prior-quarter *SUE* distribution. Based on 84,792 announcements from 1974 to 1986. *CARs* are the sums over pre- and postannouncement holding periods (beginning day −59 and day 1, respectively) of the difference between daily returns and returns for NYSE/AMEX firms of the same size decile. *SUE* represents forecast errors from a first-order autoregressive earnings expectation model (in seasonal differences) scaled by its estimation-period standard deviation (see section 3.2 for details).

Note that the immediate price response to the earnings announcement is captured by the left panel (day -1 to day 0)

The run-up in the stock price from day -60 to day -1 is either due to insider trading, or more likely, due to post-earnings announcement drift from the earnings surprise in the previous quarter (remember that the deciles are formed based on SUE's in that same period); also, remember the serial correlation in earnings surprises

The "FOS control portfolio SUE strategy" (long on the highest SUE decile and short on the lowest SUE decile) yields an estimated return of approximately 4.2 percent over the 60 days subsequent to the earnings announcement, or about 18 percent annualized

Similarly, the "continuously balanced SUE strategy" yields an annualized return of 17 percent.

The following two figures (authors' Figures 3 and 4) indicate how the drift varies by firm size

- As noted by Foster, Olsen, and Shevlin (1984), the post-announcement drift is larger for smaller companies
- Among small companies (Figure 3), the "FOS control portfolio SUE strategy" yields a return of approximately 5.3 percent over the 60 trading days subsequent to the earnings announcement
  - Returns for medium-sized companies (not shown) and large companies (Figure 4) are 4.5 and 2.8 percent, respectively
  - Results based on the "continuously balanced SUE strategy" for small, medium, and large companies amount to 5.1, 4.3, and 2.8 percent, respectively
  - In a statistical test (not reported), the authors show that the inverse relation between firm size and post-earnings- announcement drift is statistically significant

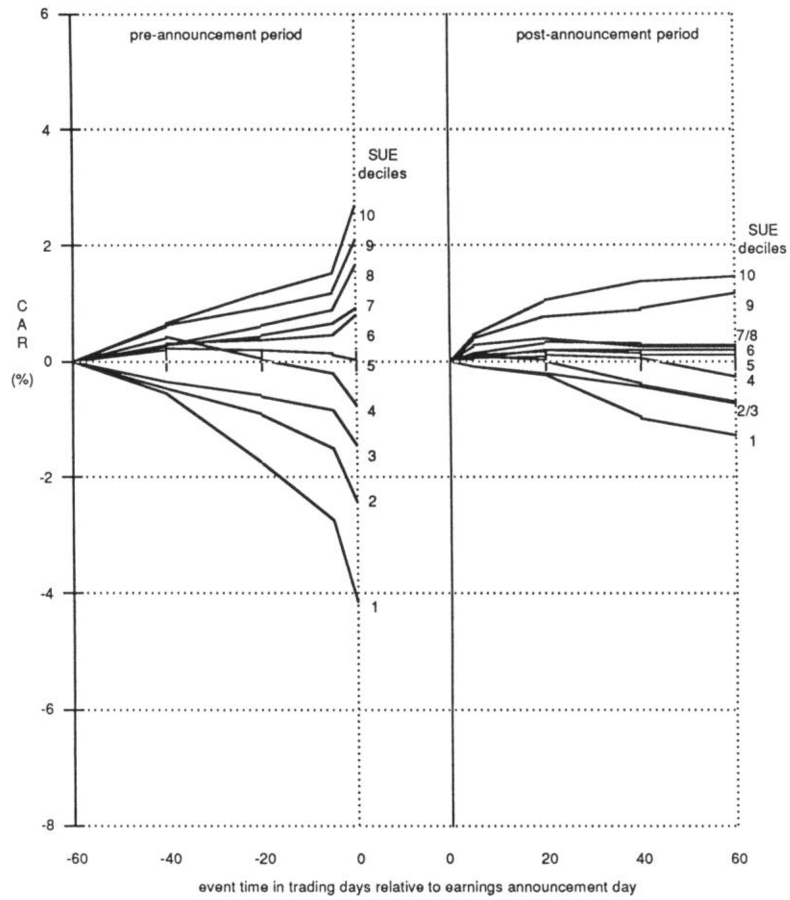


FIG. 3.—Cumulative abnormal returns (*CARs*) for *SUE* portfolios: large firms only. Earnings announcements are assigned to deciles based on standing of standardized unexpected earnings (*SUE*) relative to prior-quarter *SUE* distribution. Based on 27,584 announcements from 1974 to 1986. Large firms are in size deciles 8 to 10, based on January 1 market values of equity for all NYSE and AMEX firms. *CARs* are the sums over pre- and postannouncement holding periods (beginning day −59 and day 1, respectively) of the difference between daily returns and returns for NYSE-AMEX firms of the same size decile. *SUE* represents forecast errors from a first-order autoregressive earnings expectation model (in seasonal differences) scaled by its estimation-period standard deviation (see section 3.2 for details).

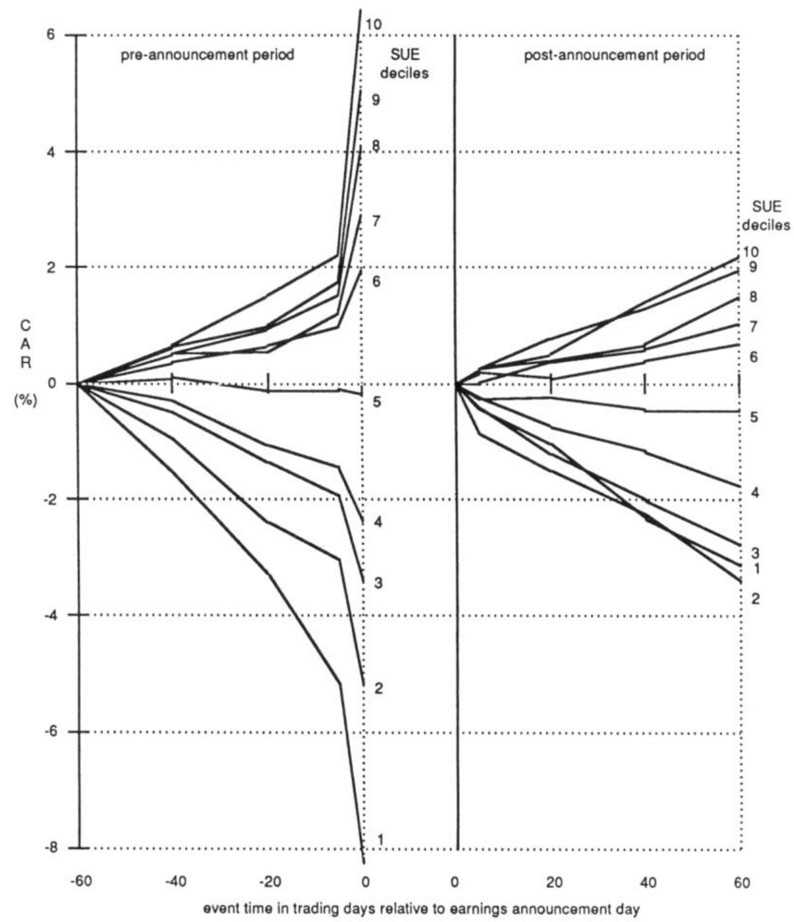


FIG. 4.—Cumulative abnormal returns (*CARs*) for *SUE* portfolios: small firms only. Earnings announcements are assigned to deciles based on standing of standardized unexpected earnings (*SUE*) relative to prior-quarter *SUE* distribution. Based on 29,796 announcements from 1974 to 1986. Small firms are in size deciles 1 to 4, based on January 1 market values of equity for all NYSE and AMEX firms. *CARs* are the sums over pre- and postannouncement holding periods (beginning day −59 and day 1, respectively) of the difference between daily returns and returns for NYSE-AMEX firms of the same size decile. *SUE* represents forecast errors from a first-order autoregressive earnings expectation model (in seasonal differences) scaled by its estimation-period standard deviation (see section 3.2 for details).

The following table (authors' Table 1) provides information about the longevity of the post-announcement drift for stocks ranked in the lowest and highest SUE decile, broken down by size and by sub-periods extending <u>two years</u> beyond the earnings announcement

> Most of the drift occurs during the first 60 days subsequent to the earnings announcement, and there is little evidence of statistically significant drift beyond 180 trading days

The percentage of the 60-day drift that occurs within 5 days amounts to 13, 18, and 20 percent for small, medium and large companies, respectively (not shown)

Essentially, all the drift occurs within nine months for small companies and within six months for large companies

Remember that when markets are efficient in the sense that the earnings information becomes fully priced as soon as it is public knowledge, the expected return on the long-short portfolio is nil

Judged by the less strict "there is no money on the table" efficiency concept, the expected return on the long-short portfolio <u>after transaction costs</u> should not exceed the risk-free rate of return if implemented as a real-world investment strategy

TABLE 1

Longevity of Post-Earnings-Announcement Drift

Cumulative abnormal returns (CAR) for high and low standardized unexpected earnings (SUE) portfolios[1]

(Those with SUE ranked in the highest and lowest deciles)

Holding Period (Trading Days, Relative to Announcement)	Small Firms			Medium Firms			Large Firms		
	High *SUE*	Low *SUE*	Diff. (Hi-Lo)	High *SUE*	Low *SUE*	Diff. (Hi-Lo)	High *SUE*	Low *SUE*	Diff. (Hi-Lo)
				Cumulative Abnormal Returns (%)					
−59 to 0	6.42*	−8.27*	14.70*	4.15*	−6.81*	10.96*	2.71*	−4.13*	6.84*
1 to 60	2.19*	−3.13*	5.32*	1.93*	−2.58*	4.51*	1.45*	−1.29*	2.74*
61 to 120	0.38	−2.24*	2.62*	0.33	−2.22*	2.55*	0.51**	−0.76*	1.27*
121 to 180	0.03	−1.93*	1.95*	0.18	−1.61*	1.79*	−0.20	−0.66*	0.45
181 to 240	0.20	−0.38	0.58	−0.40	−0.58***	0.18	−0.45**	−0.44***	−0.01
241 to 300	−1.22*	0.56	−1.77*	−0.88*	0.16	−1.04**	−0.62*	0.23	−0.85*
301 to 360	−0.54	−0.96*	0.42	−0.36	−0.71**	0.35	−0.64*	−0.16	−0.48
361 to 420	−0.27	−0.33	0.06	−0.18	−0.37	0.18	−0.55*	−0.71*	0.17
421 to 480	0.29	−0.51	0.80	−0.43	0.37	−0.80***	−0.42**	−0.73*	0.31
				Postannouncement Drift (Cumulative Abnormal Returns, %)					
1 to 60			5.32*			4.51*			2.74*
1 to 120			7.95*			7.06*			4.02*
1 to 180			9.90*			8.85*			4.47*
1 to 480			9.99*			7.72*			3.61*
				Postannouncement Drift (as a Fraction of 480-Day Drift)					
1 to 60			0.53			0.58			0.76
1 to 120			0.80			0.91			1.11
1 to 180			0.99			1.15			1.24
1 to 480			1.00			1.00			1.00

[1] *CARs* are the sums over specified holding periods of the difference between daily returns and returns for NYSE-AMEX firms of the same size decile. *SUE* represents forecast errors from a first-order autoregressive earnings expectation model (in seasonal differences) scaled by its estimation-period standard deviation (see section 3.2 for details). Small, medium, and large firms are in size deciles 1 to 4, 5 to 7, and 8 to 10, respectively, based on January 1 market value of equity for all NYSE and AMEX firms.

Significance levels for two-tailed tests of the hypotheses that abnormal returns equal zero are coded as follows:

* Significant at the 1% level.

** Significant at the 5% level.

*** Significant at the 10% level.

Tests of risk premium as explanation for post-announcement drift

The authors present results of tests designed to assess the plausibility that the return earned from post-announcement drift is compensation for risk

Note that if CAPM betas for long and short positions are equal during the post-announcement period, the "continuously balanced SUE strategy" portfolios have a beta equal to zero, i.e., the portfolios are market-neutral

The reported returns of 5.1, 4.3, and 2.8 percent for small, medium-sized, and large companies, respectively, on the "continuously balanced SUE strategy" portfolios thus are abnormal returns to the degree that they exceed the return on a risk-free asset (i.e., the risk-free rate of return).

First, the authors examine whether betas shift around the time of earnings announcements and—most importantly—whether such shifts are asymmetric between positive SUE (good news) and negative SUE (bad news) companies

The authors estimate CAPM betas for each SUE decile for pre- and post-announcement time windows and try to assess differences in these betas

For each of several 60-day windows surrounding the earnings announcement, the authors compounded total returns on individual stocks, R_{jt}, treasury bills, R_{ft}, and the value-weighted CRSP index, R_{mt}

These three data points constitute a single observation for a regression based on the Sharpe-Lintner CAPM:

$$R_{jt} - R_{ft} = a + b \cdot (R_{mt} - R_{ft}) + e_{jt}$$

Parameter a is called Jensen's alpha.

The regression equation was estimated by pooling all observations for a given SUE decile within six 60-trading-day windows surrounding the earnings announcement date

This approach permits the betas to shift from one window to the next and to vary across SUE categories.

The following table (authors' Table 2) shows the estimated CAPM betas for the different windows and SUE deciles

Note that the authors did not test for the statistical significance of the differences in betas across windows, which means that the measured differences in betas across windows contain next to no information!

TABLE 2

Beta Estimates by SUE Category, in Periods Surrounding Earnings Announcement[1]

SUE Decile (1 = low; 10 = high)	Preannouncement Period		Postannouncement Period			
	(−119, 60)	(−59, 0)	(1, 60)	(61, 120)	(121, 180)	(181, 240)
Beta estimates						
1	1.16	1.22	1.17	1.17	1.23	1.31
2	1.11	1.17	1.15	1.08	1.19	1.25
3	1.16	1.21	1.13	1.11	1.14	1.22
4	1.24	1.18	1.21	1.15	1.19	1.18
5	1.23	1.26	1.24	1.30	1.19	1.24
6	1.31	1.27	1.28	1.24	1.26	1.23
7	1.30	1.24	1.23	1.26	1.25	1.24
8	1.28	1.34	1.30	1.30	1.30	1.20
9	1.26	1.31	1.29	1.26	1.20	1.23
10	1.32	1.31	1.38	1.30	1.31	1.23
Rank correlation, *SUE* and beta	.83*	.84*	.90*	.77*	.66*	−.38
Jensen's alpha						
SUE = 1	−3.7%*	−5.3%*	−1.6%*	−0.8%*	−0.8%*	0.6%*
SUE = 10	3.4*	6.1*	3.0*	1.4*	0.7*	0.7*
Combined	7.1*	11.4*	4.6*	2.2*	1.5*	0.1

[1] For each 60-day window, we calculate compounded daily returns for individual stocks, the value-weighted *CRSP* index, and the treasury-bill rate. These data constitute a single observation in a regression of individual stock returns against market returns, both expressed in terms of differences from the treasury-bill rate. Such regressions are estimated within each *SUE* category. There are approximately 8,500 (overlapping and thus nonindependent) observations underlying estimates for the (−59, 0) and (1, 60) windows, and slightly fewer for other windows. The standard error for each estimate in the table is approximately 0.02. Cross-sectional dependence in the data may cause downward bias in the estimated standard error (Bernard [1987]).

* Significantly different from zero, .05 level (two-tailed test).

The relation between SUE decile and CAPM beta is positive and statistically significant for 60- and 120-trading-day windows around the announcement; in particular, “good news” stocks tend to have more systematic risk than “bad news” stocks during the post-announcement period

- Remember from above that the “FOS control portfolio SUE strategy” yields an estimated return of approximately 4.2 percent over the 60 trading days subsequent to the earnings announcement, or about 18 percent on an annualized basis

Similar to this result, the authors report a difference in excess return, $R_{jt} - R_{ft}$, between highest-SUE (SUE 10) and lowest-SUE (SUE 1) companies of 4.3 percent over the 60 trading days subsequent to the earnings announcement

- On the other hand, the corresponding mean excess return of the market, $R_{mt} - R_{ft}$, runs at only 1.65 percent
    - Thus, if the betas are to explain post-earnings announcement drift, the difference between betas for the SUE 10 and SUE 1 stocks would have to equal 2.6 (4.3 percent divided by 1.65 percent)
        - In fact, the difference in betas is only 0.21
- The failure of the CAPM beta to fully explain the magnitude of the drift is confirmed by the statistical (and economic significance) of Jensen’s alpha as shown in the table
    - Jensen’s alpha measures the vertical distance between the location of the stock on the (R_j, β) coordinates and the securities market line

Thus, Jensen's alpha is a measure for abnormal return, i.e., return above and beyond what is warranted by the systematic risk incorporated in stock *j*

For the "FOS control portfolio SUE strategy" the authors estimate a (statistically significant) abnormal return (i.e., difference in Jensen's alphas) of 4.6 percent (or 18 percent annualized).

Second, in the authors' effort to determine whether post-earnings announcement drift is nothing other than a risk premium (rather than an abnormal return), the authors estimate Stephen Ross' APT model

Whereas the CAPM begins with an analysis of how investors construct efficient portfolios, Ross' arbitrage pricing theory (a somewhat misleading term) does not ask which portfolios are efficient

Rather, the APT approach simply assumes that each stock's return depends partly on pervasive macroeconomic influences (or "factors") and "noise"—influences that are unique to that company

> Some of the factors might be risk factors, while others might help hedge risk (and thus have a negative influence on the asset's expected return).

The following table (authors' Table 3) presents results of an APT regression approach where the authors regress quarterly CAR (raw return minus market return, which is actually excess return) differences between SUE 10 and SUE 1 portfolios on five (macroeconomic) factors discussed in APT literature

TABLE 3

Sensitivity of Drift to Risk Factors Used in Studies of Arbitrage-Pricing Theory[1]
(Returns based on FOS control portfolio SUE strategy)

	Independent Variables								
	Intercept	$R_{mt}-R_{ft}$	*QP*	*DEI*	*UI*	*UPR*	*UTS*	*R*-Square	*F*-Test[2] of Significance of Variables Other Than ($R_{mt}-R_{ft}$)
Sign of coefficient, if risky (as opposed to a hedge)		+	+	−	−	+	−		
Coefficient	.04	.04	−.19	−.57	.42	−.08	.03	.07	.75
T-value	(8.63)	(.60)	(−.97)	(−.37)	(.44)	(−.38)	(.03)		
Coefficient	.04	−	−.16	−.92	.37	−.01	.06	.06	.58
T-value	(8.70)		(−.84)	(−.66)	(.39)	(−.04)	(1.07)		

[1] Dependent variable is the calendar-quarter return on a zero-investment portfolio, where long (short) positions in extreme good news stocks (extreme bad news stocks) are offset by positions in a portfolio of NYSE-AMEX firms in the same size decile. The returns are regressed against the return on the value-weighted NYSE index (in risk premium form) and five factors identified by Chen, Roll, and Ross [1986] as potentially influencing asset prices.

Extreme good (bad) news is defined in terms of standardized unexpected earnings (*SUE*), relative to prior-quarter *SUE* distribution. Firms ranked in the highest (lowest) decile are considered extreme good (bad) news firms.

Independent variables are defined as follows:

$R_{mt}-R_{ft}$ = return on value-weighted NYSE index, less 90-day treasury-bill rate;
QP = quarterly growth rate in industrial production, lagged ahead one period;
DEI = change in expected inflation;
UI = unanticipated inflation;
UPR = unanticipated change in the default risk premium (return on high-yield bonds [under BBB], less return on AAA bonds);
UTS = unanticipated change in the term structure (return on long-term government bonds, less treasury-bill rate).

[2] $F(5, 43)$ is significant at the .05 level for values in excess of 2.44.

The table shows that the hypothesis that the five risk factors—as a group—do not explain the return difference between the SUE 10 and SUE 1 portfolios cannot be rejected (*F*-test, insignficant, rightmost column)

At the same time, the intercept (next to leftmost column) is statistically significant, suggesting an abnormal return of 4 percent.

How consistent is the profitability of a long-short strategy on SUE 10 and SUE 1 portfolios?

- The following chart (authors' Figure 5) shows CAR's for the "FOS control portfolio SUE strategy" and the "continuously balanced SUE strategy," by calendar quarter
  - CAR's are assigned to calendar quarters based on the portion of the 60-trading-day CAR generated within that calendar quarter

The return in Panel A is positive in 46 of 50 quarters and in 13 of 13 years, while in Panel B, the returns are positive in 44 of 50 quarters and in 13 of 13 years; no catastrophic loss occured

In conclusion, the risk-return profiles of the two (self-financing) long-short portfolios (Panels A and B) are hard to reconcile with the view that these returns are a risk premium (rather than abnormal returns)

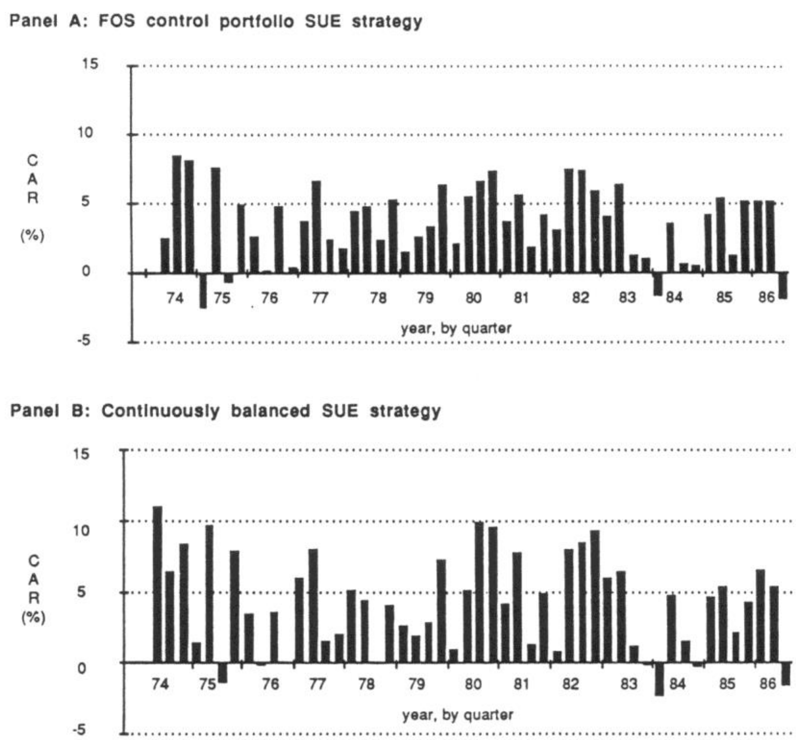


FIG. 5.—Cumulative abnormal returns (*CARs*) from *SUE* strategies, by calendar quarter. In both panels, long (short) positions are assumed in the highest (lowest) quintiles of standardized unexpected earnings (*SUE*) and held for 60 trading days. *CARs* are assigned to calendar quarters based on the portion of the 60-day *CAR* generated within that calendar quarter. *SUE* represents forecast errors from a first-order autoregressive earnings expectation model (in seasonal differences) scaled by its estimation-period standard deviation (see section 3.2 for details). In panel A, *CARs* are the combined abnormal returns from a long position in the highest *SUE* quintile and a short position in the lowest *SUE* quintile. Abnormal returns are the sums over the 60 trading days after the announcement of the difference between daily returns and returns for NYSE-AMEX firms of the same size decile. In panel B, continuous balancing requires that each $1 long position in the highest *SUE* quintile is always offset by a short position in similar-sized stocks (small, medium, or large) in the lowest *SUE* quintile. Balancing in this way sometimes requires waiting after earnings announcements until an offsetting "match" is available. *CARs*, computed over the 60 trading days after matching, are a combination of the compounded (buy and hold) returns for the long and short positions.

Finally, the authors try assess whether the post-earnings- announcement drift is caused by transaction costs

- Transaction costs might prevent traders from incorporating earnings information quickly into prices or, in other words, prevent investors from exploiting seemingly profitable long-short strategies that work toward eliminating the mispricing
  - When transaction costs are defined to include both bid-ask spreads and commissions, they are about 4 percent and 2 percent for small and large stocks, respectively
  - To calculate the cost of a long-short strategy, one must double these amounts to reflect both legs (i.e., long and short)
  - Then, taking into account that such a long-short strategy involves (on average) a 78 percent turnover each quarter, the implied cost would be about 6 and 3 percent per quarter for small and large stocks, respectively
  - These amounts are approximately equal to the 60-trading-day post-announcement abnormal returns shown in the authors' Table 1 (above)

In trying to assess the validity of the transaction cost argument, the authors investigate whether the transaction costs put an upper bound on the post-announcement drift of long-short positions with extreme SUE differences

- To this end, the authors divide the sample based on the companies' SUE rankings into halves, then thirds, quintiles, deciles (as they have done before), and so on, until finally they divide the sample into 100 portfolios
- At each of these steps, the authors calculate the post-earnings-announcement 60-trading-day CAR on a portfolio that is long on the highest SUE portfolio and short on the lowest SUE portfolio
  - Note that at each step, the SUE difference between the two portfolios becomes more extreme
  - If trading costs cause post-announcement drift, then the CAR on a long-short portfolio should be limited to these costs, no matter how large the SUE difference between the two portfolios actually is.

Panel A of the following chart (authors' Figure 6) shows the 60-trading-day CAR (vertical axis) for long-short portfolios for ever more extreme SUE differences (horizontal axis)—brought about by dividing the sample into ever finer SUE portfolios

> The graph for the post-announcement drift (which is displayed alongside with a ± one-standard-deviation interval) supports the view that there is a bound of about 4 percent (2 percent per leg) up to which SUE differences are eliminated.

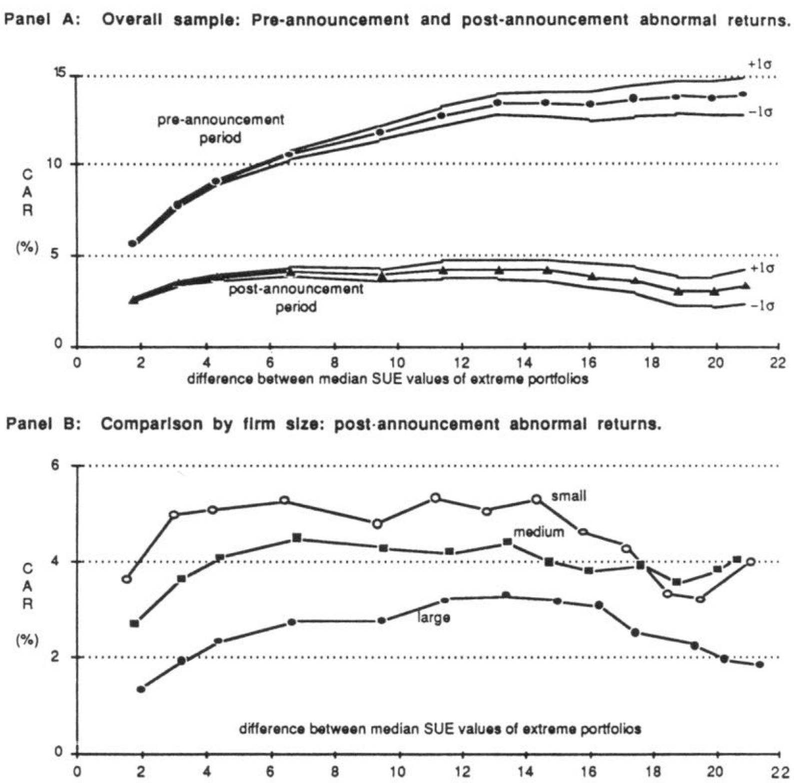


FIG. 6.—Test of an explanation for the drift, based on costs that impede trading. The plot presents the difference in drifts or cumulative abnormal returns (*CARs*) over 60 days after earnings announcements, between the most positive and most negative *SUE* (standardized unexpected earnings) portfolios, constructed by splitting the sample into 2, 3, 5, 10, 20, . . ., 100 portfolios based on *SUE*. The hypothesis predicts that, if the drift is caused by costs that impede trading, the postannouncement drift should remain less than those costs, regardless of the *SUE* difference between extreme portfolios. Thus, as differences between *SUEs* of extreme portfolios increase (represented by movement toward the right of the graph), the postannouncement *CARs* should level out, despite increases in the preannouncement *CARs*. *CARs* are the sums over 60-trading-day pre- and postannouncement holding periods of the difference between daily returns and returns for NYSE-AMEX firms of the same size decile. *SUE* represents forecast errors from a first-order autoregressive earnings expectation model (in seasonal differences) scaled by its estimation-period standard deviation (see section 3.2 for details). Small, medium, and large firms are in size deciles 1 to 4, 5 to 7, and 8 to 10, respectively, based on January 1 market values of equity for all NYSE and AMEX firms.

The upper bound is detailed in Panel B, which breaks down the sample into small, medium-size and large companies

- Remember that trading costs vary inversely with firm size
- The upper bounds for the different size classes displayed in Panel B lie within the transactions costs that have been determined in empirical studies for the size classes and the sample period in question.

Conclusion

The stock market appears to be slow in incorporating earnings information in stock prices, which generates post-earnings-announcement drift

Thus, the stock market is not efficient in the sense that all publicly available information is priced at any time (semi-strong form of market efficiency; see the chapter "Efficient Markets").

On the other hand, there is no money on the table

After accounting for transaction costs, it is hard to make money on long-short portfolios that try to arbitrage across assets that are mispriced relative to each other based on earnings information.

Least-cost dealing

From a least-cost dealing perspective, relatively overpriced stocks should be the first to be sold; conversely, new money arriving at the stock market should flow into the relatively underpriced asset

This way, least-cost dealing helps eliminate asset mispricing

On one hand, least-cost dealing is more powerful than arbitrage because transactions costs are lower (no two legs)

On the other hand, least-cost dealing relies on investors liquidating and buying anyway (and when doing this, looking for the best deal).

Empirical Studies

15. Predictability of Corporate Earnings

Reference:

Chan, Louis K. C., Jason Karceski and Josef Lakonishok (2001) “The Level and Persistence of Growth Rates,” NBER Working Paper No. 8282, http://www.nber.org.

Note: All tables in this chapter are from Chan, Karceski and Lakonishok (2001)

Abstract of Chan, Karceski and Lakonishok (2001)

> “Expected long-term earnings growth rates are crucial inputs to valuation models and for cost of capital estimates. We analyze historical long-term growth rates across a broad cross-section of stocks using several operating performance indicators. We test whether growth persists, and whether it is [predictable]. Cases of very high growth have occurred, but are relatively rare. There is scant persistence in growth beyond chance, and limited ability to identify companies with high future long-term growth. Regressions using a variety of predictors confirm the low predictability of growth. Valuations that assume persistently high growth over prolonged periods rest on shaky foundations.”

Note that no company can have an earnings growth rate of more than 3.5 percent annual forever, inflation-adjusted

- The long-term rate of real GDP growth runs at 3.5 percent and the share of corporate profits in national income is limited (and mean-reverting)
- Consequently, so-called growth stocks (which are frequently glamour stocks) cannot be growth stocks forever
  - Eugene F. Fama and Kenneth R. French (2001, “Forecasting Profitability and Earnings,” *Journal of Business* 73(2), 161-175) estimates that <u>profitability</u> (ratio of earnings before interest and extraordinary items but after taxes, divided by the book value of assets) <u>mean-reverts at a rate of 38 percent annual</u>
  - Hence, you might want to be cautious before buying into a “growth stock.”

Chan, Karceski and Lakonishok (2001), henceforth “the authors,” do not explicitly account for mean-reversion of earnings growth in their study on the predictability of earnings (of individual companies)

- (Remember that mean-reversion implies predictability, at least to some degree)
- The authors’ contribution is to show that …
  - … earnings growth shows little persistence
  - … identifying the next Microsoft (i.e., the next company with a multi-year run of greatly above-average earnings growth) is a game of chance
  - … analysts do not add much to the predictability of earnings growth beyond the information subsumed in the company’s dividend yield.

Analysts show great confidence in predicting corporate earnings

(Remember that the unconditional mean should be the forecast when there is no predictability)

For instance, forward P/E ratios (forward price-to-earnings multiples) vary greatly in the IBES (http://www.firstcall.com, operated by Thomson Financial, a widely followed provider of corporate earnings forecasts) universe of U.S. companies

At year-end 1999, the distribution of the ratio of stock price to analyst consensus forecasts of the following year's earnings has a 90th-percentile of 53.9 while the 10th-percentile is 7.4

Companies with a record of sustained, strong past growth in earnings are heavily represented among those rewarded with rich multiples (P/E ratios or P/B [price-to-book] ratios, which correlate strongly; see the chapter "Predictability of Returns").

Similarly, at year-end 1999 the distribution of the IBES five-year forecasts for average annual earnings growth has a 90th-percentile of 40 while the 10th-percentile is 8.9

Remember from the chapter "Heuristic and Biases" that …

… people tend to ignore the baseline frequency of outcomes in the presence of noise (representativeness bias)

… people deliver overly narrow confidence intervals (anchoring).

Hence, it is not surprising to discover a gap between share valuation along with analysts' predictions on one hand, and realized growth of operating performance on the other

For instance, take a company with a forward P/E ratio of 100

(At the stock market peak in March 2000, the Nasdaq 100 or more than ten percent of the U.S. market capitalization traded in this neighborhood)

Assume that the P/E ratio reverts to 20 within 10 years—a conservative measure given the historical average of around 15—and the return on the stock over this 10-year period is zero (no capital gain, no dividends)

In such a scenario, the earnings of this company would have to grow by 17.5 percent a year!

Alternatively, if the investors demanded a capital gain of 10 percent (no dividends), the earnings would have to grow by an annual 29.2 percent

Casual evidence suggests that the return investors expect on glamour stocks—many of which trade at P/E ratios around 100—greatly exceeds 10 percent.

The authors' sample comprises all domestic stocks with data on the Compustat Active and Research files

Companies are selected at the end of each calendar year from 1951 to 1997

The number of eligible companies grows from 359 in the first sample selection year to 6,825 in the last year; on average the sample comprises about 2,900 companies.

Note that <u>none</u> of the numbers in this study is <u>inflation-adjusted</u>.

The authors look at three measures of operating performance

Net sales

Operating income before depreciation

Income before extraordinary items

This frequently used measure is beset with pitfalls

Often times, this number is negative; consequently, the growth rate is undefined

Recovering beaten-down companies and startups that start to generate profits often exhibit very high growth rates, which distorts the picture.

In the analysis, the authors take the perspective of an investor who buys and holds the stock over some time period

All numbers are on a per-share basis, adjusted for dividends and stock splits

Dividends (and special contributions, such as spin-offs) are reinvested to put companies with different payout policies on equal footing.

Distribution of historical growth rates over long horizons (five and ten years)

- Percentiles are calculated for the distribution obtained at each year-end, and then the percentiles are averaged across years
  - Note that because dividends (and other distributions) are reinvested, the calculated growth rates are higher than conventional growth rates
  - Also, note the survivorship bias (in another section of the paper the authors document an upward bias in growth rates due to survivorship)
    - Not all companies survive the time period in question
    - For negative base values (not applicable for sales), no growth rates are calculated.
  - To illustrate the survivorship problem, on average there are about 2,900 companies available for inclusion at each year-end
    - Of these, 2,782 companies survive until the end of the next year and have a reported value for income before extraordinary items
      - Of these companies, 1,994 have positive values for income in the base year, i.e., can be used to calculate growth rates.
    - At the five-year horizon, there are on average 1,884 surviving companies, for 1,398 of which growth rates can be calculated.
    - At the ten-year horizon, there are on average 1,265 surviving companies; growth rates are calculated for 1,002 of these, while the remainder has negative base values.

- Table 1 provides a sobering reality check for analysts and investors who flock to stocks with rich price-earnings multiples
  - Take the aforementioned example of a stock with a price-earnings ratio of 100 that declines to 20 in ten years' time with an expected return of 10 percent a year
    - As mentioned, earnings must grow at 29.2 percent per year over ten years to justify the current multiple
      - This is a tall order by historical standards: the growth rate corresponds to about the 95th-percentile of the distribution of ten-year growth rates (with dividends reinvested).
    - Suppose earnings grow at a historically more representative—but still healthy—annual rate of 14.7 percent (the 75th-percentile)
      - In this case, the current P/E ratio of 100 would be justified if the time for the multiple to fall to 20 is stretched out to 38 years!

Table 1
Distribution of growth rates of operating performance
over 1, 5 and 10 years: All firms

At every calendar year-end over the sample period growth rates in operating performance are calculated over each of the following one, five and ten years for all firms in the sample. The sample period is 1951–1998, and the sample includes all domestic firms listed on the New York, American and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample selection date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. Percentiles of the distribution are calculated each year-end; the simple average over the entire sample period of the percentiles is reported, along with the distribution of growth rates over horizons ending in the last three years of the sample period.

Part I: Annualized growth rate over 10 years

Sample period	Percentile 2%	10%	25%	40%	50%	60%	75%	90%	98%
	(A) Sales								
Average	-9.6	0.1	5.5	8.7	10.2	11.5	13.8	18.0	27.6
Ending 1996	-20.5	-3.7	3.0	6.7	8.4	10.4	13.6	20.2	36.2
Ending 1997	-21.0	-3.6	3.1	6.6	8.4	10.3	13.5	20.8	36.8
Ending 1998	-16.1	-3.4	2.9	6.2	7.9	9.5	12.7	19.2	32.9
	(B) Operating income before depreciation								
Average	-13.3	-2.3	4.1	7.6	9.5	11.2	14.1	19.4	31.3
Ending 1996	-16.7	-3.8	3.4	7.6	9.5	11.4	15.3	22.5	39.2
Ending 1997	-16.8	-3.2	3.6	7.6	9.3	11.4	15.1	23.9	41.9
Ending 1998	-14.6	-3.3	3.3	7.2	9.0	10.9	14.1	21.5	38.6
	(C) Income before extraordinary items								
Average	-15.6	-3.1	3.9	7.7	9.7	11.6	14.7	20.4	33.4
Ending 1996	-17.2	-4.2	3.6	8.1	10.2	12.4	16.7	26.6	48.1
Ending 1997	-18.2	-3.6	3.7	8.1	10.3	12.6	17.6	27.3	48.5
Ending 1998	-21.2	-6.3	2.3	6.9	9.0	11.4	15.3	24.4	48.8

Part II: Annualized growth rate over 5 years

Sample period	Percentile 2%	10%	25%	40%	50%	60%	75%	90%	98%
	(A): Sales								
Average	-18.7	-4.1	4.3	8.2	10.2	12.0	15.3	22.1	40.5
Ending 1996	-32.3	-9.9	0.4	5.5	7.9	10.1	14.9	25.8	52.1
Ending 1997	-30.4	-8.3	1.3	6.6	9.2	11.4	16.4	27.9	64.2
Ending 1998	-22.7	-6.2	2.9	8.0	10.2	12.4	17.1	27.6	56.3
	(B) Operating income before depreciation								
Average	-26.8	-8.4	1.9	7.2	9.8	12.4	17.1	26.7	51.2
Ending 1996	-30.8	-9.3	2.1	8.4	11.3	14.2	20.4	34.9	73.8
Ending 1997	-31.3	-9.9	3.0	9.1	11.9	14.8	21.8	35.2	71.7
Ending 1998	-24.4	-7.8	3.5	8.7	11.5	14.4	19.9	33.4	64.4
	(C) Income before extraordinary items								
Average	-30.9	-10.3	1.5	7.4	10.5	13.4	18.8	30.4	62.4
Ending 1996	-35.1	-10.5	3.2	10.2	13.8	17.4	26.9	47.7	108.1
Ending 1997	-36.1	-10.4	3.6	9.9	13.2	16.8	25.8	45.5	92.5
Ending 1998	-35.1	-11.5	2.8	9.1	12.4	15.7	23.1	40.1	88.2

Part III: 1-year growth rate

Sample period	Percentile 2%	10%	25%	40%	50%	60%	75%	90%	98%
	(A) Sales								
Average	-47.3	-12.9	1.2	7.6	10.9	14.2	21.0	38.7	121.7
Ending 1996	-61.0	-20.2	-1.8	6.4	10.5	14.6	24.8	50.0	176.2
Ending 1997	-60.4	-20.8	-1.0	7.0	11.0	15.6	26.1	57.8	204.6
Ending 1998	-58.3	-20.8	-1.4	6.3	10.3	14.5	24.9	54.1	181.9
	(B) Operating income before depreciation								
Average	-69.4	-30.7	-5.6	5.9	11.8	17.7	30.6	67.4	253.3
Ending 1996	-74.1	-30.8	-2.6	9.0	14.7	21.3	36.7	88.7	334.3
Ending 1997	-77.6	-31.0	-0.9	9.8	15.2	21.4	37.0	83.2	314.8
Ending 1998	-74.1	-34.7	-4.9	6.7	12.2	18.5	32.2	76.5	273.2
	(C) Income before extraordinary items								
Average	-76.8	-37.9	-7.4	6.9	13.3	19.9	35.8	90.2	435.3
Ending 1996	-87.8	-46.8	-9.5	9.6	17.4	25.5	47.7	140.2	720.8
Ending 1997	-88.0	-47.1	-6.4	11.4	19.2	28.0	53.1	137.0	631.0
Ending 1998	-87.3	-48.2	-13.7	5.4	13.7	21.3	40.4	115.0	727.2

Table 2 reports the distribution of growth rates for large companies only (companies in the two top deciles of year-end equity market capitalization)

- Bigger companies have a larger scale of operations and hence are more likely to face limits on their growth
  - Hence, extremely high growth rates are less prevalent in Table 2 compared to Table 1
    - For example, the 90th-percentiles of growth over 10 years for all three measures of operating performance are close to 16 percent a year
      - Note that dividend yields are higher for large companies, i.e., the numbers calculated for large firms overestimate the actual growth rates by a wider margin than for small companies.

Table 2

Distribution of growth rates of operating performance over 1, 5 and 10 years: Large firms

At every calendar year-end over the sample period growth rates in operating performance are calculated over each of the following one, five and ten years for large firms (in the top two deciles of year-end equity market capitalization, based on NYSE breakpoints). The sample period is 1951–1998, and the sample includes all domestic firms listed on the New York, American and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. Percentiles of the distribution are calculated each year-end; the simple average over the entire sample period of the percentiles is reported, along with the distribution of growth rates over horizons ending in the last three years of the sample period.

Part I: Annualized growth rate over 10 years

Sample period	Percentile 2%	10%	25%	40%	50%	60%	75%	90%	98%
(A): Sales									
Average	-3.4	2.5	6.8	9.4	10.7	11.7	13.3	16.3	22.0
Ending 1996	-4.6	1.0	6.2	8.7	9.6	10.5	12.3	16.3	26.2
Ending 1997	-7.5	1.5	5.8	8.4	9.7	10.5	12.1	16.2	24.8
Ending 1998	-7.7	-0.2	4.4	6.7	8.5	9.5	11.1	15.0	21.5
(B) Operating income before depreciation									
Average	-8.3	0.6	5.4	8.1	9.5	10.8	12.9	16.1	22.6
Ending 1996	-5.3	2.9	7.0	9.2	10.3	11.4	14.2	19.0	34.5
Ending 1997	-10.5	2.3	6.8	8.8	9.6	11.0	13.4	18.4	33.6
Ending 1998	-11.6	-1.7	4.3	7.4	8.7	10.4	11.8	16.3	21.4
(C) Income before extraordinary items									
Average	-12.8	-0.9	4.5	7.5	9.3	10.8	13.1	16.6	23.8
Ending 1996	-8.8	-1.1	5.9	8.7	10.6	12.3	15.0	21.4	49.9
Ending 1997	-22.6	-2.8	4.0	7.3	9.2	11.4	14.5	23.5	43.8
Ending 1998	-25.6	-3.8	1.7	6.1	8.2	9.9	13.3	18.5	36.4

Part II: Annualized growth rate over 5 years

Sample period	Percentile 2%	10%	25%	40%	50%	60%	75%	90%	98%
(A): Sales									
Average	-9.7	-0.6	6.9	9.4	10.8	11.9	14.1	18.1	27.9
Ending 1996	-11.3	-3.4	4.3	7.3	8.4	9.9	12.3	16.7	31.2
Ending 1997	-10.8	-1.8	5.2	8.3	9.5	10.8	13.2	18.3	30.1
Ending 1998	-13.6	-3.0	4.0	8.8	10.2	11.5	13.7	19.6	32.5
(B) Operating income before depreciation									
Average	-16.9	-3.5	4.3	7.9	9.8	11.5	14.3	19.3	32.1
Ending 1996	-14.0	-1.8	5.9	9.2	11.2	12.4	15.8	22.7	45.2
Ending 1997	-10.4	-1.5	6.6	9.6	11.0	12.7	15.7	22.4	42.4
Ending 1998	-13.6	-6.6	4.5	7.5	10.8	12.7	15.6	19.9	32.0
(C) Income before extraordinary items									
Average	-26.4	-6.4	2.8	7.6	9.8	12.0	15.3	21.3	37.2
Ending 1996	-18.9	-3.9	3.4	10.5	12.7	14.8	20.0	42.8	89.3
Ending 1997	-32.8	-6.9	2.5	8.5	11.7	14.2	18.6	28.0	53.2
Ending 1998	-39.5	-10.1	4.3	9.5	11.8	14.4	19.6	30.4	57.4

Part III: 1-year growth rate

Sample period	Percentile 2%	10%	25%	40%	50%	60%	75%	90%	98%
(A): Sales									
Average	-36.4	-2.4	5.7	9.3	11.3	13.3	17.0	25.2	47.7
Ending 1996	-46.9	-9.3	3.6	8.4	11.2	13.5	17.5	29.3	68.5
Ending 1997	-42.7	-11.5	2.3	7.5	10.3	13.3	18.6	35.5	65.4
Ending 1998	-49.8	-14.7	1.5	6.6	8.9	11.8	18.1	29.1	53.0
(B) Operating income before depreciation									
Average	-52.3	-15.2	0.2	7.1	10.6	13.8	19.8	33.7	82.3
Ending 1996	-58.8	-17.5	1.5	8.3	12.2	15.2	21.5	34.5	69.9
Ending 1997	-44.8	-20.1	0.1	8.4	11.4	14.5	19.8	37.8	104.1
Ending 1998	-60.0	-30.3	-1.9	6.6	11.1	14.0	20.8	33.4	73.1
(C) Income before extraordinary items									
Average	-67.5	-25.3	-2.8	6.9	11.0	14.9	23.1	45.9	216.6
Ending 1996	-81.3	-38.1	-6.8	10.0	15.7	18.5	31.2	95.4	395.0
Ending 1997	-88.0	-44.2	-11.7	5.4	11.7	16.3	26.1	61.0	196.6
Ending 1998	-80.0	-46.9	-13.5	4.7	11.5	15.5	27.1	56.7	213.6

In summary, the estimated median growth rate is reasonable compared to the overall economy's growth rate

On average, over the sample period, the median growth rate over ten years for income before extraordinary items is about 10 percent for all companies

The behavior over the last three (overlapping) ten-year periods in the sample roughly matches the overall average (see Tables 1 and 2)

Growth in the other two measures of operating performance also exhibits comparative medians.

After subtracting the dividend yield (the median yield is 2.5 percent) as well as inflation (which averages 4 percent per year over the sample period), the growth in real income before extraordinary items is roughly 3.5 percent per year

This number is consistent with the historical growth rate of real GDP, which has averaged 3.4 percent per year over 1950-98

It is difficult to see how—over the long haul—profitability of the business sector can grow much faster than the overall gross domestic product.

Persistence in growth

It typically takes more than one or two years of high growth to ignite investors' enthusiasm for a stock

Rather, many high-flying stocks have a track record of consistently superior growth over several years

Conversely, stocks that have done poorly over prolonged periods are shunned and trade at low multiples.

The difference in valuation based on the <u>growth record</u> indicates a <u>pervasive belief that stocks with high or low future growth are easily identifiable ex ante</u>

For instance, analysts and investors seem to believe that a company whose past growth puts it in the top tier of growth rates for several years in a row is highly likely to repeat this performance in the future.

The authors define persistence in growth rates as achieving a growth rate above the median for a consecutive number of years (called "runs")

In order to go around the problem of undefined rates of growth when base year values are negative, the authors <u>assign</u> growth rates to companies in years with negative base value according to the following principle

The authors rank the companies by their change in earnings normalized by the share price (rather than the base-year earnings number); companies with negative base-year earnings numbers are assigned the arithmetic mean of the earnings growth rates of the neighboring companies (given that they are valid).

At each year-end over the sample period the authors calculate how many firms achieve runs over horizons of one to ten years in the future

Note that because survivors are likely to perform better than the population, survivors have a greater chance of being above the median.

Table 3 reports runs, averaged across year-ends

- For growth in sales (panel A), for instance, out of an average number of 2,900 companies available for sample selection at year-end, 2,771 companies on average survive until the end of the following year
  - Over the following ten years there are, on average, 1,265 surviving companies
    - Of these companies, 11 have sales growth rates that exceed the median in each of the ten years, representing about 0.87 percent of the eligible 1,265 survivors
    - If sales growth is independent across time periods, we should expect to see 0.5^{10} (about 0.098 percent)
    - Hence, there seems to be some persistence in sales growth rates, the survivorship bias notwithstanding.
  - Over a five-year horizon, on average 118 companies, or 6.3 percent of the 1,878 companies who exist over the full five years, turn in runs above the median
    - The number expected under the hypothesis of independence over time is about 59 (or 3.1 percent of 1,878)
    - Again, it is fair to conclude that there is a certain degree of persistence in sales growth rates, the survivorship bias notwithstanding.

Panel B of Table 3 shows that there is much less (if any) persistence in operating income before deprecation

- On average, 67 companies a year, or 3.6 percent of 1,833 surviving companies, have above-median runs for <u>five</u> consecutive years, compared to 57 companies that can be expected by pure chance
- On average, 4 companies a year, or 0.3 percent of 1,223 survivors, have above-median runs for <u>ten</u> years in a row, which is 3 more than expected by chance
- Note that the respective number for the most recent ten-year periods do not differ markedly from the historical average.

Virtually <u>no persistence</u> can be detected in Panel C (of Table 3), which looks at income before extraordinary items

- An average of 57 companies (3 percent) beat the median for five years in a row, while 59 (3.1 percent) are expected to do so
- Runs above the median for ten years occur in 0.2 percent of 1,265 cases (or 2 companies), roughly matching the expected frequency (0.1 percent, or 1 company).

Table 3
Persistence in growth rates of operating performance: All firms

At every calendar year-end over the sample period growth rates in operating performance are calculated over each of the following one to ten years (or until delisting) for all firms in the sample. The sample period is 1951–1998, and the sample includes all domestic firms listed on the New York, American and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales (panel A), operating income before depreciation (panel B), or income before extraordinary items available to common equity (panel C). Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. For each of the following ten years the number of firms with valid growth rates; the number of firms whose growth rate exceeds the median growth rate each year for the indicated number of years; the percentage these firms represent relative to the number of valid firms; and the percentage expected under the hypothesis of independence across years, are reported. Statistics are provided for the entire sample period, and for the ten-year horizons corresponding to the last three sample formation years, 1987–1989.

	Firms with above-median growth each year for number of years:									
Variable	1	2	3	4	5	6	7	8	9	10
(A): Sales										
Average number of valid firms	2771	2500	2263	2058	1878	1722	1590	1471	1364	1265
Average number above median	1386	721	382	209	118	70	42	26	17	11
Percent above median	50.0	28.8	16.9	10.2	6.3	4.0	2.7	1.8	1.3	0.9
1987–1996	50.0	29.2	17.5	11.6	7.9	5.5	3.8	2.7	1.8	1.3
1988–1997	50.0	29.1	17.9	11.6	7.8	5.4	3.9	2.4	1.7	1.2
1989–1998	50.0	30.0	18.6	11.9	7.8	5.6	3.4	2.4	1.5	1.2
(B) Operating income before depreciation										
Average number of valid firms	2730	2456	2219	2014	1833	1678	1546	1428	1322	1223
Average number above median	1365	628	290	136	67	34	18	10	6	4
Percent above median	50.0	25.6	13.0	6.8	3.6	2.0	1.2	0.7	0.5	0.3
1987–1996	50.0	25.5	13.1	7.5	4.5	2.7	1.7	1.0	0.7	0.5
1988–1997	50.0	25.2	13.1	7.1	4.0	2.3	1.3	1.0	0.6	0.4
1989–1998	50.0	25.0	13.1	7.0	4.0	2.1	1.3	0.8	0.5	0.5
(C) Income before extraordinary items										
Average number of valid firms	2782	2509	2271	2065	1884	1727	1593	1473	1365	1265
Average number above median	1391	625	277	125	57	28	14	7	4	2
Percent above median	50.0	24.9	12.2	6.0	3.0	1.6	0.9	0.5	0.3	0.2
1987–1996	50.0	24.7	12.1	6.7	3.8	1.9	1.0	0.6	0.3	0.2
1988–1997	50.0	24.1	12.1	6.1	2.7	1.3	0.7	0.4	0.3	0.1
1989–1998	50.0	24.8	12.2	5.7	2.6	1.3	0.8	0.5	0.2	0.0
Expected percent above median	50.0	25.0	12.5	6.3	3.1	1.6	0.8	0.4	0.2	0.1

Table 4 repeats the exercise of Table 3 for different categories of stocks, as defined by the authors

- The median value the companies have to beat is the median value of the entire sample (not just the one that applies to the specific category of stock)
- Technology stocks are defined by SIC codes (238; 357; 366; 38; 48; 737); these stocks are R&D intensive and belong predominantly to the computer equipment, software, electrical equipment, communications and pharmaceutical industries
  - Sales show a certain degree of persistence for technology stocks (as it does for stocks overall [Table 3] and for all other categories of stocks displayed in Table 4)
  - There is also a great deal of persistence in the rate of growth of operating income, but little difference to the expected chance frequency for income before extraordinary items
    - The finding of persistence in growth of operating income might be due to the survivorship bias, which might be particularly acute for technology companies.
- Value stocks are stocks that rank in the three lowest deciles by market-to-book value
- Glamour stocks are stocks that comprise the three highest deciles by market-to-book value
  - The popular sentiment is that persistence in growth extends to glamour stocks generally
    - Yet, when glamour stocks are compared with value stocks, the two measures of earnings growth are nearly identical for the two categories of stocks (and close to the chance frequencies).

Large caps rank in the top two deciles by market capitalization, while small caps rank in the bottom three deciles, and midcaps in the remaining five deciles

While sales growth tends to be more persistent for large companies, it does not translate into persistent growth in the bottom-line income numbers.

Table 4

Persistence in growth rates of operating performance: Selected equity classes

At every calendar year-end over the sample period growth rates in operating performance are calculated over each of the following one to ten years (or until delisting) for all firms in the sample. The sample period is 1951–1998, and the underlying sample includes all domestic firms listed on the New York, American and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. For each of the following ten years the number of firms whose growth rate exceeds the median growth rate each year for the indicated number of years is expressed as a percentage of the number of firms with valid growth rates. Statistics are provided for the following sets of stocks: technology stocks (panel A), comprising stocks whose SIC codes begin with 283, 357, 366, 38, 48, or 737; value stocks (panel B), comprising stocks ranked in the top three deciles by book-to-market value of equity; glamour stocks (panel C), comprising an equivalent number as in panel B of the lowest-ranked stocks by book-to-market value of equity; large stocks (panel D), comprising stocks ranked in the top 2 deciles by equity market value; mid-cap stocks (panel E), comprising stocks ranked in the third through seventh deciles by equity market value; and small stocks (panel F), comprising stocks ranked in the bottom three deciles by equity market value. All decile breakpoints are based on domestic NYSE stocks only.

Variable	Percent of firms with above-median growth each year for number of years:									
	1	2	3	4	5	6	7	8	9	10
	(A) Technology stocks									
Sales	51.6	30.7	19.1	12.5	8.5	5.9	4.2	3.0	2.3	1.7
Operating income	51.0	27.2	14.9	8.7	5.3	3.3	2.2	1.4	1.0	0.7
Income before extraordinary items	50.9	25.9	13.5	7.3	4.1	2.5	1.5	0.9	0.5	0.4
	(B): Value stocks									
Sales	50.6	30.0	18.2	11.1	6.9	4.3	2.8	1.9	1.3	0.9
Operating income	49.3	25.3	13.2	6.8	3.5	1.8	0.9	0.5	0.3	0.2
Income before extraordinary items	48.3	23.8	11.4	5.4	2.5	1.2	0.7	0.4	0.3	0.2
	(C) Glamour stocks									
Sales	48.3	26.6	15.1	8.5	4.7	2.7	1.7	1.0	0.8	0.6
Operating income	50.1	25.2	11.9	5.9	3.3	1.7	1.0	0.6	0.4	0.3
Income before extraordinary items	50.7	25.2	12.0	5.8	2.9	1.6	0.9	0.4	0.2	0.1
	(D) Large stocks									
Sales	53.2	31.3	18.9	11.7	7.5	4.8	3.2	2.2	1.6	1.1
Operating income	49.4	25.2	13.0	6.9	3.7	2.0	1.1	0.6	0.4	0.3
Income before extraordinary items	46.7	21.9	10.0	4.7	2.2	1.2	0.7	0.4	0.3	0.2
	(E) Mid-cap stocks									
Sales	53.9	32.4	19.8	12.1	7.6	4.9	3.3	2.2	1.5	1.0
Operating income	50.5	26.6	13.9	7.5	4.2	2.4	1.5	1.0	0.7	0.4
Income before extraordinary items	49.4	24.9	12.4	6.2	3.1	1.6	0.9	0.5	0.3	0.2
	(E) Small stocks									
Sales	47.0	26.1	14.7	8.6	5.2	3.2	2.1	1.4	1.0	0.7
Operating income	50.1	25.2	12.6	6.4	3.3	1.8	1.0	0.6	0.4	0.2
Income before extraordinary items	51.0	25.5	12.6	6.3	3.2	1.7	0.9	0.4	0.2	0.1
Expected percent above median	50.0	25.0	12.5	6.3	3.1	1.6	0.8	0.4	0.2	0.1

In summary, analysts and investors seem to believe that companies' earnings grow at high rates for several years

The evidence suggests that the number of such instances is not much different from what might result from mere chance

The lack of persistence in earnings growth agrees with the notion that in competitive markets abnormal profits tend to dissipate over time.

Comparison with IBES consensus forecasts

- Historically, some companies have enjoyed torrid growth rates in excess of thirty percent a year for prolonged periods
  - If such companies are identifiable ex ante, then price-to-earnings ratios in excess of a hundred might not be unwarranted
  - Stock analysts are not shy about making aggressive growth forecasts so they are apparently confident in their ability to pick the future success stories
    - Note that the dispersion between the top and bottom deciles of IBES long-term forecasts is about 31 percentage points!
- For the following analysis (and for the remainder of the authors' study), rates of earnings growth are calculated without dividends reinvested
  - Note that IBES forecasts, which refer to income before extraordinary items, do not assume dividends being reinvested.
- In Table 8 the authors compare realized rates of earnings growth (rates of growth of <u>income before extraordinary items</u>) with IBES <u>long-term</u> consensus (median) earnings forecasts, which refer to a <u>three- to five-year horizon</u>
  - The sample period begins in 1982 (and runs through 1998) because IBES forecasts are not available prior to 1982
  - At each year-end, the authors rank all domestic companies with available IBES long-term forecasts and sort them into quintiles (20-percent quantiles)
  - The table tracks for each sample selection date (each year-end) the growth rates of companies that survive over the next one, three or five years in each quintile
    - The median realized growth rate over the companies in each quintile is then averaged across all sample selection dates.

Table 8 shows a large dispersion in IBES long-term consensus forecasts, with analysts boldly distinguishing between companies with high and low growth prospects

- The median estimate in quintile 1 averages 6 percent, while the median estimate in quintile 5 is 22.4 percent, the difference amounting to a staggering 16.6 percentage points (bottom panel)
- On the other hand, the dispersion in the realized five-year growth rates amounts to only 7.5 percentage points (panel D).

Notably, analysts' estimates are quite optimistic

- For all quintiles, the IBES growth forecasts (bottom panel) vastly exceed the realized rates of growth (panel D).

On the other hand, the IBES long-term median forecasts (bottom panel), match up fairly well with the realized growth rates over the one-year horizon (panel A).

Table 8
Realized median growth rates of operating performance for stocks classified by IBES long-term growth forecasts

At every calendar year-end t over the sample period stocks are ranked and classified to one of five groups based on IBES forecasts of long-term earnings growth. Results are reported for individual stocks and for portfolios. For individual stocks, growth rates in operating performance are calculated over each of the five subsequent years (years $t+1$ to $t+5$) for all firms in the sample with available data. The sample period is 1982–1998, and all domestic firms listed on the New York, American and Nasdaq markets with data on the Compustat files are eligible. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends. The median realized growth over all stocks in each classification is calculated each year, and the simple average over the entire sample period is reported. For portfolios, a value-weighted portfolio is formed at each year-end from all the stocks in each quintile sorted by IBES forecasts. The portfolio's income before extraordinary items is calculated over each of the subsequent five years, with the proceeds from liquidating delisted stocks reinvested in the surviving stocks. Growth rates for each portfolio are calculated in each formation year, and the simple average over the entire sample period of the growth rates is reported. Also reported are the ratios of: the prior year's income before extraordinary items per share to current price; and the prior year's cumulative regular dividends per share to current price.

	Quintile based on IBES forecast:				
Growth in:	1 (Low)	2	3	4	5 (High)
(A): Growth rate in year $t+1$					
Sales	1.4	4.5	6.3	8.3	13.7
Operating income before depreciation	3.6	6.8	7.6	10.3	16.0
Income before extraordinary items	5.1	9.5	10.1	12.0	18.3
Portfolio income before extraordinary items	12.6	4.2	4.5	7.2	13.6
No. with positive base & survive 1 year	242	256	266	318	584
No. with negative base & survive 1 year	71	78	60	88	265
(B): Growth rate in year $t+2$					
Sales	1.7	4.5	6.4	7.8	11.6
Operating income before depreciation	3.2	7.0	8.4	9.9	14.0
Income before extraordinary items	4.7	9.9	10.5	12.2	16.4
Portfolio income before extraordinary items	6.9	7.5	6.1	9.1	10.6
No. with positive base & survive 2 years	225	235	244	296	497
No. with negative base & survive 2 years	62	75	59	85	252
(C): Annualized growth rate over 3 years					
Sales	1.1	4.0	5.6	7.3	11.3
Operating income before depreciation	2.5	5.2	6.8	8.1	10.9
Income before extraordinary items	3.1	7.4	7.0	9.0	11.5
Portfolio income before extraordinary items	9.0	7.3	5.2	7.1	11.4
No. with positive base & survive 3 years	202	209	230	263	439
No. with negative base & survive 3 years	67	70	56	82	217
(D): Annualized growth rate over 5 years					
Sales	1.2	3.4	5.1	6.9	9.9
Operating income before depreciation	2.2	5.1	6.8	7.3	9.2
Income before extraordinary items	2.0	6.5	6.5	8.0	9.5
Portfolio income before extraordinary items	8.0	10.7	7.2	7.7	11.3
No. with positive base & survive 5 years	182	179	201	233	356
No. with negative base & survive 5 years	57	63	50	68	170
Median IBES forecast	6.0	10.2	12.3	15.1	22.4
Median stock dividend yield, %	6.0	3.4	2.7	1.5	0.1
Portfolio dividend yield, %	6.9	4.6	3.3	2.5	1.3
Median stock earnings to price ratio, %	10.0	8.9	7.9	7.2	5.6

Do analysts add value?

The authors entertain a horse race between highly paid stock analysts and the stock’s dividend yield (last four [quarterly] dividend payments divided by current share price) in predicting operating performance

Companies might retain earnings for investment and growth, or pay them out as dividends

In equilibrium, at the margin the shareholder is indifferent between keeping the dividend invested in the company and having it paid out (and invested elsewhere)

High dividend yields come at the expense of future growth

Generally, in equilibrium, a company’s dividend policy signals the availability of positive-NPV (net present value) projects.

In summary, the dividend yield is expected to have predictive power for the earnings growth rate

Generally, in equilibrium, a company’s dividend policy signals the availability of positive-NPV (net present value) projects.

The authors’ estimate the following regression model:

At every calendar year-end a cross-sectional regression model is used to forecast growth rates of operating performance, y_{it+j}, for firm i over the following one to five years for all firms in the sample with available data. The model is

$$y_{it+j} = \beta_0 + \beta_1 PASTGS5_{it} + \beta_2 TECH_{it} + \beta_3 BM_{it} + \beta_4 PASTR6_{it} + \beta_5 IBESLTG_{it} + \beta_6 DP_{it} + \epsilon_{it+j}$$

The dependent variable is growth in: sales (SALES); operating income before depreciation (OIBD); or income before extraordinary items available to common equity (IBEI). The variables used to forecast a firm’s growth are: $PASTGS5$, the growth in sales over the five years prior to the sample selection date; $TECH$, a dummy variable with a value of one for a stock in the technology sector and zero otherwise; BM, book-to-market ratio; $PASTR6$, the stock’s prior six-month compound rate of return; $IBESLTG$ the IBES consensus forecast for long-term growth; and DP the dividend yield, accumulated regular dividends per share over the last twelve months divided by current price per share.

The authors use, alternatively, all three measures of operating performance as dependent variables

The forecast equation incorporates explanatory variables that are popularly thought to connote high growth

For instance, there is an indicator variable for companies from the technology and pharmaceutical sectors (*TECH*).

To be eligible, a company must have no missing value for the time horizon in question and a positive base-year value of the operating performance variable in question (otherwise the growth rate is not defined)

The displayed R^2 is in fact the $\bar{R}^2$ (adjusted R^2)

Although on shaky theoretical ground, an increase in the $\bar{R}^2$ in response to adding an explanatory variable in a regression equation to an existing set of explanatory variables if often read as this variable having additional explanatory power, that is having predictive power beyond the information contained in the existing set of explanatory variables

Note that the sign of the dividend yield in the regression equation is expected to be negative, while the sign of the IBES long-term forecast is expected to be positive.

Panel D shows the regression equation for the out-of-sample prediction of long-term (five-year) growth

- IBES long-term forecasts are statistically significant and have the expected sign
- When the dividend yield is added, …
  - … the regression coefficient has the expected sign and is statistically significant (save for income before extraordinary income)
  - … the $\bar{R}^2$ increases (which, roughly, can be read as the dividend yield adding predictive power)
  - … the *t*-statistics of the IBES regression coefficient decreases.
- Judged by values of the $\bar{R}^2$ statistics, the predictability of long-term growth seems to be poor.

Unfortunately, the horse-race the authors entertained is not state of the art and allows no conclusion about the relative predictive power of IBES forecast vis-à-vis the predictive power of the dividend yield

- To compare the predictive power of the long-term IBES forecast with the predictive power of the dividend yield, the authors would have to conduct a *J*-test, which is a set of two tests
  - The *J*-test, which is not always conclusive, tries to assess which variable drives out which (in terms of explanatory power).

Table 9

Forecasting regressions for growth rates of operating performance

At every calendar year-end a cross-sectional regression model is used to forecast growth rates of operating performance, y_{it+j}, for firm i over the following one to five years for all firms in the sample with available data. The model is

$$y_{it+j} = \beta_0 + \beta_1 PASTGS5_{it} + \beta_2 TECH_{it} + \beta_3 BM_{it} + \beta_4 PASTR6_{it} + \beta_5 IBESLTG_{it} + \beta_6 DP_{it} + \epsilon_{it+j}$$

The dependent variable is growth in: sales (SALES); operating income before depreciation (OIBD); or income before extraordinary items available to common equity (IBEI). The variables used to forecast a firm's growth are: $PASTGS5$, the growth in sales over the five years prior to the sample selection date; $TECH$, a dummy variable with a value of one for a stock in the technology sector and zero otherwise; BM, book-to-market ratio; $PASTR6$, the stock's prior six-month compound rate of return; $IBESLTG$ the IBES consensus forecast for long-term growth; and DP the dividend yield, accumulated regular dividends per share over the last twelve months divided by current price per share.

Growth in:	PASTGS5	TECH	BM	PASTR6	IBESLTG	DP	$\bar{R}^2$
			(A): Growth rate in year $t+1$				
SALES	0.1212	-0.0019	-0.0198	0.0535	0.2986	-0.3935	0.0592
	(5.2)	(-0.3)	(-5.6)	(4.3)	(6.2)	(-4.4)	
OIBD	-0.1031	0.0071	0.0005	-0.0859	0.2812	-0.7036	0.0167
	(-1.7)	(0.6)	(0.1)	(-4.0)	(3.1)	(-4.8)	
IBEI	-0.1412	-0.0105	0.0076	-0.0929	0.1679	-1.1380	0.0135
	(-1.8)	(-0.5)	(0.4)	(-3.6)	(1.2)	(-3.9)	
			(B): Growth rate in year $t+2$				
SALES	0.0672	0.0015	-0.0241	0.0432	0.2156	-0.3709	0.0398
	(2.4)	(0.3)	(-6.0)	(4.6)	(7.5)	(-3.4)	
OIBD	-0.0828	0.0094	0.0096	-0.0382	0.3156	-0.5417	0.0116
	(-1.4)	(1.0)	(1.7)	(-1.5)	(5.0)	(-2.3)	
IBEI	0.0029	-0.0174	0.0204	-0.0908	0.1427	-0.6077	0.0115
	(0.0)	(-1.2)	(1.5)	(-1.9)	(1.1)	(-1.5)	
			(C): Annualized growth rate over years $t+1$ to $t+3$				
SALES	0.0813	0.0040	-0.0253	0.0444	0.1858	-0.4370	0.0827
	(2.4)	(1.3)	(-8.6)	(5.2)	(9.5)	(-7.3)	
OIBD	-0.0513	0.0030	-0.0085	-0.0164	0.1350	-0.4247	0.0161
	(-1.2)	(0.6)	(-1.7)	(-2.2)	(2.8)	(-2.3)	
IBEI	0.0014	-0.0021	-0.0089	-0.0430	0.1148	-0.1428	0.0121
	(0.1)	(-0.2)	(-0.8)	(-5.4)	(1.4)	(-0.6)	
			(D): Annualized growth rate over years $t+1$ to $t+5$				
SALES	0.0635	0.0071	-0.0272	0.0274	0.2315		0.0908
	(2.1)	(2.1)	(-7.7)	(4.2)	(5.6)		
OIBD	-0.0578	0.0045	-0.0095	-0.0217	0.1947		0.0211
	(-2.0)	(1.9)	(-2.3)	(-4.3)	(2.8)		
IBEI	-0.0227	-0.0014	-0.0130	-0.0426	0.1284		0.0191
	(-3.1)	(-0.2)	(-1.2)	(-4.1)	(2.0)		
SALES	0.0513	0.0045	-0.0265	0.0342	0.1508	-0.4397	0.1005
	(1.7)	(1.3)	(-9.1)	(4.4)	(2.9)	(-18.6)	
OIBD	-0.0667	0.0035	-0.0090	-0.0189	0.1387	-0.2969	0.0242
	(-2.3)	(1.4)	(-2.5)	(-3.2)	(1.8)	(-5.2)	
IBEI	-0.0246	-0.0012	-0.0126	-0.0411	0.0966	-0.1455	0.0206
	(-2.4)	(-0.2)	(-1.2)	(-4.4)	(1.3)	(-1.1)	

Growth in each operating peformance variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends. $PASTGS5$, $PASTR6$ are Winsorized at their 5-th and 95-th percentiles; $IBESLTG$ is Winsorized at its 1-st and 99-th percentiles; and DP is Winsorized at its 98-th percentile. Stocks with negative values of BM are excluded. In the regressions for $OIBD$ or

[Blank page]

Theoretical Models

16. Noise Trader Risk (Limits of Arbitrage I)

References:

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann (1990) "Noise Trader Risk in Financial Markets," *Journal of Political Economy* 98, 703-38.

Hakansson, Nils H., and William T. Ziemba (1995) "Capital Growth Theory," in: Robert A. Jarrow, Vojislav Maksimovic, and William T. Ziemba, ed., *Finance*, Handbooks in Operations Research and Management Science, Vol. 9, Ch. 3, Amsterdam: Elsevier, 65-86.

Shleifer, Andrei (2000) Inefficient Markets: An Introduction to Behavioral Finance. Oxford: Oxford University Press, Ch. 2.

Arbitrage

- Arbitrage in its purest form is defined as the simultaneous purchase and sale of the same security in two different markets for advantageously different prices
  - Arbitrage plays a critical role in securities markets because it works toward the elimination of security mispricing
  - Theoretically, arbitrage requires no capital and entails no risk
  - Future cash flows of arbitrage transactions are zero and mispricing is eliminated instantly.
  - Real-world arbitrage rarely fits the above textbook description
    - Mispricing rarely occurs between two identical securities (and if it does, arbitrage will eliminate it quickly); rather, mispricing happens between similar securities (as the following example of shares of Royal Dutch Shell illustrates)
    - The risk of arbitrage is that the securities mispricing might deepen in the short run
    - Institutions that attract outside money to engage aggressively on risk arbitrage (correction of securities mispricing) are called hedge funds (a misnomer).

Noise traders

A major reason for securities mispricing is the existence of noise traders

Noise traders, as opposed to sophisticated investors, are unable to distinguish information from noise

Noise traders trade on noise as if it were information

Noise traders provide liquidity to financial markets

There is empirical evidence that stock markets are more volatile when open than when closed

This empirical finding indicates that there is more trading than the arrival of new information warrants.

Noise traders cause securities prices to deviate from intrinsic value

The intrinsic (or fundamental) value of a security is the "true" value of the security given the state of information about its future cash flows at the time

Securities mispricing offers arbitrage opportunities

Arbitrage is risky in that noise traders might deepen the securities mispricing in the short run (noise trader risk).

Modeling noise traders

Early asset pricing models with noise traders concluded that there is no sustained securities mispricing

Noise trader risk was treated as idiosyncratic, which means that it could be eliminated through diversification

Risk that can be eliminated is not priced.

Note that noise traders trade more with informed traders than they trade with other noise traders, because the noise traders and the informed traders are eager to bet against each other

In early noise trader models in which noise traders have no sustained influence on securities prices, noise traders (on average) suffer losses when trading with informed investors and are driven out of the market

Note that there always are some noise traders around, because new noise traders enter the market and existing noise traders keep investing (part of) their labor income.

There is reason to believe that noise trading may lead to <u>sustained</u> deviations of securities prices from their intrinsic values

- If the time horizon of arbitrageurs is finite (rather than infinite), <u>arbitrage is limited</u> in its ability to eliminate mispricing
  - Most arbitrageurs are agents for institutions or wealthy investors
    - Investors evaluate arbitrageurs at regular, relatively short intervals and pay them according to their performance
      - The longer an arbitrageur keeps losing money (despite being on the right side of the market), the more likely the principal is inclined to believe that the arbitrageur is on the wrong side of the market.
  - Many arbitrageurs borrow securities (when short-selling) and money from intermediaries to put on their trades
    - Arbitrageurs face a liquidity risk because they have to make interest payments on loans, and they might face margin calls if the value of the collateral falls
    - There is also a risk of having to return borrowed securities if the owners of these securities decide to sell.

The “twin-securities” case of Royal Dutch and Shell Transport

Froot, K.A., and E. Dabora (1999) “How are Stock Prices Affected by the Location of Trade,” *Journal of Financial Economics* 53, 189-216

Royal Dutch and Shell are independently incorporated in the Netherlands and England, respectively

For details on the company history see <http://www.shell.com/>

The Royal Dutch/Shell group emerged from an alliance between Royal Dutch and Shell Transport and Trading in 1907

The companies merged their interests on a 60:40 ratio; all cash flows are effectively split in these proportions.

Royal Dutch trades primarily in the Netherlands and the United States (and is included in the S&P 500)

Shell trades predominantly in the United Kingdom (and is included in the Financial Times Allshare Index, FTSE)

If the market values of securities were equal to the net present values of future cash flows, the value of Royal Dutch equity should be equal to 1.5 times the value of Shell equity

As the following figure shows, there are enormous deviations from parity, ranging from the relative underpricing of Royal Dutch by 35 percent to relative overpricing by 10 percent

Remember that both kinds of stocks are about equally liquid, which means that a possible liquidity premium cannot explain the price differences

In a market where arbitrageurs have infinite time horizon and face no transaction costs, an arbitrageur can simply go long on the relatively underpriced shares and go short on the relatively overpriced shares, and hold the hedged position—if necessary—forever

FIG. 2.1 Log deviations from Royal Dutch/Shell parity.
Source: Froot and Dabora (1998).

Source: Shleifer (2000, p. 30)

The above figure illustrates the risk of mispricing deepening

An arbitrageur who went long on the relatively underpriced Royal Dutch shares and went short on the relatively overpriced Shell shares in mid 1983 when the mispricing was 10 percent, experienced a severe deterioration of his position as the mispricing widened to nearly 25 percent six months later

- If this arbitrageur was leveraged or had to explain his losses to investors, he might have failed to survive through these losses and might have been forced to liquidate this position
- Note that it took about four years for the 30 percent mispricing that prevailed in September 1980 to go away.

Another example of a violation of the law of price are the prices of stocks of 3Com and its subsidiary Palm in the year 2000

- See the chapter "Valuation of 3Com and Palm"
  - Owen and Thaler argue that the price pressure from short-selling was limited through a shortage of stocks available for borrowing.

A model of noise trading

The following theoretical model shows that the unpredictability of investor sentiment—of future noise trader demand for assets—becomes an important source of risk and a deterrent to arbitrage

The model shows that the conclusion of unsophisticated investors earning lower average returns than the arbitrageurs is unwarranted

The model

There are arbitrageurs (informed investors) and noise traders

All noise traders are identical, and so are all arbitrageurs

Noise traders form erroneous beliefs about the future distribution of returns on a risky asset

Note that because all noise traders are identical, the erroneous beliefs are common to all noise traders.

It is optimal for arbitrageurs to exploit noise traders' misperceptions

The model will show that the arbitrageurs' investment strategies push prices toward fundamentals, but not all the way.

The model is a simple overlapping generations (OG) model

A new generation of agents is born every period

All agents live for two periods

There are two types of agents

There are noise traders (denoted n) that are present in measure μ, and there are arbitrageurs (denoted a) that are present in measure $1-\mu$

There is no first-period consumption, no labor supply or bequest

The resources agents have available to invest are exogenous

The only decision agents make is to choose a portfolio when young (i.e., at the beginning of the first period).

There are two assets that make identical dividend (or coupon) payments

The safe asset, s, pays a fixed real (i.e., inflation-adjusted) dividend, r, per period

A unit of the safe asset can be created out of, and a unit of the safe asset can be turned into, a unit of the consumption good in any period

With consumption each period taken as numeraire, the price of the safe asset is always fixed at one

The dividend r paid on asset s is the risk-free rate.

The risky asset, u, always pays the same fixed real dividend r as asset s

Unlike asset s, asset u is not in elastic supply: its quantity is fixed and equal to one

The price of u in period t equals p_t

If the price of asset u were equal to the present value of its dividends (i.e., its intrinsic value), then assets u and s would be perfect substitutes and would sell at the same price in all periods

The model shows that there might be sustained deviations of the price the risky asset, u, from intrinsic value.

The model assumes limited risk-bearing capacity of arbitrageurs

The limit on the arbitrageur's risk-bearing capacity may result from market-wide noise trader risk or from noise trader risk that affects a large group of stocks

The bubble in telecom stocks in the late 1990s that started deflating in March 2000, is an example for such correlation

When a large fraction of the market (e.g., Nasdaq 100) is affected by noise trader risk, arbitrageurs cannot diversify, which limits the aggressiveness of their trades.

Both types of agents choose their portfolios when young (in period t) to maximize perceived expected utility given their own beliefs about the ex ante mean of the distribution of the price of u at $t+1$

(Remember that the expected utility is the expected value of the utility levels that are associated with the respective possible states of nature)

The representative arbitrageur (young in period t) accurately perceives the distribution of returns from holding the risky asset

The representative noise trader (young in period t) misperceives the expected price of the risky asset by an independent and identically distributed normal random variable ρ_t:

$$\rho_t \sim N(\rho^*, \sigma_\rho^2)$$

The mean misperception ρ^* is a measure of the average "bullishness" of noise traders (as measured across generations; note that within every generation, all noise traders are identical)

The results that are derived in the following depend critically on the unpredictability of investor sentiment.

The constant σ_ρ^2 is the variance of noise traders' misperceptions of the expected return per unit of risky asset

Noise traders maximize their own expectation of utility given the next-period dividend, the one-period variance of p_{t+1}, and their false belief that the distribution of the price of u next period has mean ρ_t above intrinsic value.

Each agent's utility is a constant absolute risk aversion (CARA) function of "final" wealth, that is, wealth when he is old:

$$U = -e^{-(2\gamma)w}$$

where 2γ is the coefficient of absolute risk aversion and w is final wealth

Remember the definition of absolute risk aversion, r_a:

$$r_a = -\frac{U''(w)}{U'(w)}$$

Given their beliefs, all young agents divide their portfolios between u and s

When old, agents convert their holdings of s into the consumption good, sell their holdings of u for the price p_{t+1} to the young, and consume all their wealth

With normally distributed returns to holding a unit of the risky asset, maximizing a CARA function is equivalent to maximizing the following mean-variance utility:

$$w - \gamma \cdot \sigma_w^2$$

where w is the expected final wealth, and σ_w^2 is the one-period ahead variance of wealth (and 2γ is the Arrow-Pratt degree of absolute risk aversion)

The <u>representative sophisticated investor</u> chooses the amount λ_t^a of the risky asset u to maximize

$$\mathrm{E}(U) = w - \gamma \cdot \sigma_w^2$$

$$= \lambda_t^a \cdot [r + {}_t p_{t+1} - p_t (1+r)] - \gamma \cdot (\lambda_t^a)^2 ({}_t \sigma_{p_{t+1}}^2)$$

The anterior subscript denotes the time at which an expectation is taken

The term $({}_t \sigma_{p_{t+1}}^2)$ is the one-period variance of p_{t+1}:

$$({}_t \sigma_{p_{t+1}}^2) = \mathrm{E}_t \{[p_{t+1} - E_t (p_{t+1})]^2\}$$

The <u>representative noise trader</u> chooses the amount λ_t^n of the risky asset u to maximize

$$\mathrm{E}(U) = w - \gamma \cdot \sigma_w^2$$

$$= \lambda_t^n \cdot [r + {}_t p_{t+1} + \rho_t - p_t (1+r)] - \gamma \cdot (\lambda_t^n)^2 ({}_t \sigma_{p_{t+1}}^2)$$

The only difference between the objective functions of the arbitrageur and the noise trader is the term $\lambda_t^n \cdot \rho_t$, which captures the noise traders' misperceptions of the expected return from holding λ_t^n units of the risky asset

Maximizing the objective functions with respect to the quantities of the risky asset held by the two representative individuals, λ_t^a, λ_t^n, yields the following two demand functions:

$$\lambda_t^a = \frac{r + {}_t p_{t+1} - p_t (1+r)}{2\gamma \cdot ({}_t \sigma_{p_{t+1}}^2)}$$

$$\lambda_t^n = \frac{r + {}_t p_{t+1} + \rho_t - p_t (1+r)}{2\gamma \cdot ({}_t \sigma_{p_{t+1}}^2)}$$

Arbitrageurs' and noise traders' demands for the risky asset are allowed to be negative (although not for both parties of the same generation), which means that they are allowed to take short positions

The fact that returns are unbounded gives each party the chance of having negative final wealth (although not for both parties of the same generation)

The demands for the risky asset are proportional to its perceived excess return (i.e., return over the risk-free asset) and inversely proportional to its perceived variance

The only difference between the arbitrageur's and the noise trader's demand function is the term $\frac{\rho_t}{2\gamma \cdot ({}_t\sigma^2_{p_{t+1}})}$, which captures the noise traders' misperceptions of the expected return of the risky asset

When noise traders overestimate expected returns, they demand more of the risky asset than the arbitrageurs; if they underestimate the expected return, they demand less

Arbitrageurs exert a stabilizing influence in this model since they offset the volatile positions of the noise traders

This finding casts doubt on the destabilizing role of hedge funds (a misnomer for "risk arbitrage funds") in financial markets, a much-touted hypothesis by politicians (who actually know better) and people susceptible to conspiracy theories.

The variance of prices in the denominators of the demand functions is derived solely from noise trader risk

Both noise traders and arbitrageurs limit their demand for the risky asset because the price at which they can sell it once they are old depends on the uncertain beliefs of the next period's young noise traders

This uncertainty about the price at which asset u can be sold afflicts both parties and prevents them from taking unlimited bets against each other

This way, noise traders keep arbitrageurs from driving prices all the way to fundamental values.

To calculate equilibrium prices, observe that the old sell their holdings to the young, which means that the demand of the young for the risky asset (which is in fixed supply) must sum up to unity:

$$(1-\mu)\cdot\lambda_t^a + \mu\cdot\lambda_t^n = 1$$

After inserting the respective demand functions we obtain for the price of the risky asset, u, in period t, p_t:

$$p_t = \frac{1}{1+r}[r + ({}_t p_{t+1}) - 2\cdot\gamma\cdot({}_t\sigma^2_{p_{t+1}}) + \mu\cdot\rho_t]$$

The price of the risky asset, p_t, is a function of …

… misperception by noise traders (ρ_t) in period t

… technological (r) and behavioral (γ) parameters

… the moments (expected value; variance) of the one-period-ahead distribution of p_{t+1}

We consider only steady-state equilibria by imposing the restriction that the unconditional distribution of p_{t+1} be identical to the distribution of p_t

The endogenous one-period-ahead distribution of the price of asset u can thus be eliminated from the above pricing function by solving recursively:

$$p_t = 1 + \frac{\mu \cdot (\rho_t - \rho^*)}{1+r} + \frac{\mu \cdot \rho^*}{r} - \frac{2 \cdot \gamma}{r} \cdot ({}_t\sigma^2_{p_{t+1}})$$

Technically, this recursive solution is arrived at by substituting ρ^*_t ($\equiv \mathrm{E}_t[\rho_t]$) for the random variable ρ_t, substituting p_t for p_{t+1}, and then solving for p_t

The solution reads (after rearranging terms):

$$p_t = 1 + \frac{\mu \cdot (\rho_t - \rho^*)}{1+r} + \frac{\mu \cdot \rho^*}{r} - \frac{2 \cdot \gamma}{r} \cdot ({}_t\sigma^2_{p_{t+1}})$$

Only the second term, $\frac{\mu \cdot (\rho_t - \rho^*)}{1+r}$, is variable (random due to the randomness of investor sentiment, ρ_t), for γ, ρ^*, and r are all constants, and the one-period-ahead variance of p_t is a simple unchanging function of the constant variance of a generation of noise traders' misperceptions ρ_t:

$${}_t\sigma^2_{p_{t+1}} = \sigma^2_{p_{t+1}} = \frac{\mu^2 \sigma^2_\rho}{(1+r)^2}$$

The final form of the price of u, which depends only on exogenous parameters of the model and on public information about present and future misperceptions by noise traders, reads:

$$p_t = 1 + \frac{\mu \cdot (\rho_t - \rho^*)}{1+r} + \frac{\mu \cdot \rho^*}{r} - 2 \cdot \gamma \cdot \frac{\mu^2 \sigma^2_\rho}{r \cdot (1+r)^2}$$

The last three terms of the price of asset u show the impact of noise traders (remember that if there were no noise traders the price of the risky asset would be unity)

- The second term, $\frac{\mu \cdot (\rho_t - \rho^*)}{1+r}$, captures the fluctuations in the price of the risky asset u due to the variation in noise traders' misperceptions
  - Even though asset u is not subject to any fundamental risk and is known to be so by a large class of investors (the arbitrageurs), its price varies substantially as noise traders' opinions change
    - When a generation of noise traders is more bullish than the average generation, they bid up the price
    - Conversely, when a generation of noise traders is more bearish than the average generation, they bid down the price
    - When noise traders hold their average misperception, the term is zero.
- The third term, $\frac{\mu \cdot \rho^*}{r}$, captures the deviations of p_t from its fundamental value due to the fact that the average misperception by noise traders [note that within every generation, all noise traders are identical], ρ^*, is not zero
  - If noise traders are bullish on average, this "price pressure" effect makes the price of the risky asset higher than it would otherwise be.

The fourth term, $-2 \cdot \gamma \cdot \frac{\mu^2 \sigma_\rho^2}{r \cdot (1+r)^2}$, captures the effect of noise trader risk on the price of the risky asset

Both parties (arbitrageurs; noise traders) present in period t believe that the asset is mispriced (but the parties think it is mispriced in different directions)

Because p_{t+1} is uncertain, neither party is willing to bet too much on this mispricing

At the margin, the return from enlarging one's position is offset by the additional risk that must be borne

The uncertainty over next period's noise traders' beliefs makes the otherwise risk-free asset u risky, driving down its price and increasing its return

This way, noise traders "create their own space."

A brief discussion of the economic implications of three fundamental assumptions of the model

(1) Overlapping generations

If there were a last period, both parties would engage in unlimited bets against each other in what they perceive to be risk-free arbitrage

Each party's horizon is limited, which means that neither party has any opportunity to wait until the price of the risky asset recovers before selling

The limited time horizon corresponds to real-world institutional features such as frequent evaluations of money managers' performance

If the horizon of arbitrageurs is long relative to the duration of noise traders' optimism or pessimism toward risky assets, then they can buy low, confident of being able to sell high when prices revert to the mean

In general, the longer the horizon of the arbitrageurs, the more aggressively they trade, and the more efficient are the markets

The longer the arbitrageurs' horizon, the more periods there are in which they can liquidate the position

The longer the time horizon, the bigger is the share of dividends in expected returns (dividend payments as an insurance mechanism).

(2) Fixed supply of risky asset

This assumption prevents the arbitrageurs to convert relatively underpriced assets into relatively overpriced assets

> In the dual share example of Royal Dutch Shell, this assumption fits the data
>
> In other instances, this assumption is at odds with reality
>
> There is empirical evidence that companies (successfully) time the market when issuing stock (initial public offerings [IPOs] or seasoned equity offerings [SEOs]).

(3) Noise trader risk is systematic

Noise trader risk affects the market as a whole or a significant segment of traded securities

> If noise trader risk were purely idiosyncratic, it would not be priced (because it could be diversified away)
>
> If noise trader risk is systematic, there is co-movement across the affected securities prices.

Relative returns of noise traders and arbitrageurs

When noise trader risk is idiosyncratic, noise trader risk is not priced, and noise traders' expected value from trading with informed traders is negative

The expected return of noise traders from holding risky assets is lower than the return of the informed traders

Besides, the more noise traders trade, the worse they do

For empirical evidence, see, for instance, Terrance Odean (1999) "Do Investors Trade Too Much?" *American Economic Review* 89, 1279-89, and other papers of the same author.

However, when noise trader risk is systematic, noise traders might enjoy higher returns than informed traders

If noise traders portfolios are concentrated in assets subject to noise trader risk, noise traders can earn higher average rates of return on their portfolios than arbitrageurs

The difference between the <u>changes in wealth</u> (from initial wealth in period t to final wealth in period $t+1$) between the two groups of investors is solely due to differences in the fractions of the risky asset held by the two groups, λ_t^a and λ_t^n:

$$\Delta W_{n-a} = (\lambda_t^n - \lambda_t^a)\cdot[r + p_{t+1} - p_t \cdot (1+r)]$$

For the <u>difference in demand</u>, $\lambda_t^n - \lambda_t^a$, we can write on the basis of what we have established above:

$$\lambda_t^n - \lambda_t^a = \frac{\rho_t}{2\cdot\gamma\cdot({}_t\,\sigma^2_{p_{t+1}})} = \frac{(1+r)^2\,\rho_t}{2\cdot\gamma\cdot\mu^2\sigma_\rho^2}$$

Note that as μ (measure for [fraction of] noise traders) becomes small, noise trader risk decreases and the bets the two groups place against each other increase.

For the ex-ante excess payoff of the risky asset u (i.e., payoff in excess of the payoff of the risk-free asset s as of time t), ${}_t[r + p_{t+1} - p_t \cdot (1+r)]$, we obtain with the help of the pricing function, $p_t = \frac{1}{1+r}[r + {}_t p_{t+1} - 2 \cdot \gamma \cdot ({}_t \sigma^2_{p_{t+1}}) + \mu \cdot \rho_t]$:

$$
\begin{aligned}
{}_t[r + p_{t+1} - p_t \cdot (1+r)] &= 2 \cdot \gamma \cdot ({}_t \sigma^2_{p_{t+1}}) - \mu \cdot \rho_t \\
&= \frac{2 \cdot \gamma \cdot \mu^2 \sigma^2_\rho}{(1+r)^2} - \mu \cdot \rho_t
\end{aligned}
$$

Consequently, we can write for the ex-ante difference in the change in wealth between the two groups of investors:

$$
{}_t(\Delta W_{n-a}) = \rho_t - \frac{(1+r)^2 \rho_t^2}{2 \cdot \gamma \cdot \mu \cdot \sigma^2_\rho}
$$

We thus obtain for the expected value of the difference in changes in wealth across the two parties of investors, using the statistical property $E(\rho^2) = [E(\rho)]^2 + Var(\rho)$:

$$
E(\Delta W_{n-a}) = \rho^* - \frac{(1+r)^2 (\rho^*)^2 + (1+r)^2 \sigma^2_\rho}{2 \cdot \gamma \cdot \mu \cdot \sigma^2_\rho}
$$

The equation shows that "bullishness" (a positive ρ^*) is a necessary condition for noise traders to earn higher expected returns than arbitrageurs

The first term, ρ^*, increases noise traders' return through the "hold more" effect

The noise traders' expected return relative to the return of the arbitrageurs increases when noise traders hold more of the risky asset and earn a larger share of the rewards of risk bearing

Note that when ρ^* is negative (noise traders are bearish), the unpredictability of investor sentiment (i.e., noise traders' misperceptions) still makes asset u risky and still pushes up the expected return on asset u

The rewards to risk bearing accrue disproportionately to arbitrageurs, who on average hold more of the risky asset than the noise traders do.

The first term in the numerator, $(1+r)^2(\rho^*)^2$, incorporates the "price pressure" effect

As noise traders become more bullish, they demand more of the risky asset on average and drive up its price

The noise traders thus reduce the return on risk bearing and, hence, the difference between their returns and those of arbitrageurs.

The second term in the numerator, $(1+r)^2\ \sigma_\rho^2$, incorporates the "buy high – sell low" or Friedman effect (named after Milton Friedman who published in 1953 one of the first papers on noise traders)

As noise traders become more bullish, they demand more of the risky asset on average and drive up its price

Because noise traders' misperceptions are stochastic, they have the worst possible market timing

Noise traders buy the most of the risky asset u when other noise traders are buying it, which is when they are most likely to suffer a capital loss

Just think of all the people who invested in telecom stocks in the late 1990s.

The more variable noise traders' beliefs are, the more damage their poor market timing does to their returns.

The denominator, $2 \cdot \gamma \cdot \mu \cdot \sigma_\rho^2$, incorporates the "create space" effect

As the variability of the noise traders' beliefs increases, so does price risk

To take advantage of noise traders' misperceptions, arbitrageurs must bear this greater risk

Because arbitrageurs (like noise traders) are risk-averse, they limit their bets against the noise traders.

If the "create space" effect is large, then the "price pressure" and "buy high – sell low" effects inflict less damage on noise traders' average returns relative to arbitrageurs' returns.

Neither party clearly dominates in terms of expected returns

The "hold more" and "create space" effects tend to raise noise traders' relative expected returns ("relative" meaning "compared to arbitrageurs")

The "buy high – sell low" and the "price pressure" effects tend to lower noise traders' relative returns

Noise traders cannot earn higher than average returns if they are on average bearish [remember that within every generation, all noise traders are identical], for if the average misperception, ρ^*, does not exceed zero, there is no "hold more" effect and $E(\Delta W_{n-a}) = \rho^* - \frac{(1+r)^2 (\rho^*)^2 + (1+r)^2 \sigma_\rho^2}{2 \cdot \gamma \cdot \mu \cdot \sigma_\rho^2}$ is consequently negative

Also, noise traders cannot earn higher than average returns if they are too bullish on average [note that within any generation, all noise traders are identical], for as ρ^* gets large, the price pressure effect, which increases with $(\rho^*)^2$, dominates

For intermediate degrees of average bullishness, ρ^*, noise traders earn higher expected returns than informed investors

- Investors evaluate arbitrageurs at regular, relatively short intervals and pay them according to their performance

  - Equation $E(\Delta W_{n-a}) = \rho^* - \frac{(1+r)^2 (\rho^*)^2 + (1+r)^2 \sigma_\rho^2}{2 \cdot \gamma \cdot \mu \cdot \sigma_\rho^2}$ shows that the larger γ is, that is, the more risk-averse agents are, the larger is the range of ρ^* over which noise traders earn higher average returns

  - No blanket statement can be made that noise traders lose money and eventually become unimportant.

Noise traders are unhappy people

Since sophisticated investors maximize true (rather than misperceived) expected utility, any trading strategy alternative to theirs that earns a higher mean return must have a variance sufficiently higher to make it unattractive

Sophisticated investors are necessarily better off when noise traders are present in this model

The presence of noise traders gives sophisticated investors a larger opportunity set in that they can still invest all they want at the risk-free rate r.

Noise traders are likely to end up poor

Expected returns are not the same as survival

It can be shown (see the chapter "Judging Investment Strategies") that long run survival in financial markets relies on a delicate trade-off between risk and expected return

Among rational investors with different utility functions, those with logarithmic preferences over final wealth have the highest probability of keeping their wealth above a given level in the long run

While both types of agents have the same utility function (and thus the same degree of risk aversion, γ), high average bullishness (misperception of fundamentals) leads noise traders to (unintentionally) take on more risk than what their degree of risk aversion calls for.

Investment advice

Do not try to ride stock market bubbles

Sell stocks the public has fallen in love with and do not get back in after seeing the stock keep going up

Many of those who tried to ride the telecom bubble in the late 1990s wound up with gaping losses

There was a young lady from Niger
Who smiled as she rode on a tiger
They returned from the ride
With the lady inside
And a smile on the face of the tiger

Do not trade frequently (Don't lose money "coming and going")

Most personal investors fall into the category of noise traders, <u>and so might you</u>

Remember that when you trade, you are more likely to trade with an informed investor than with a noise traders (because noise trader risk tends to be correlated), which means that you are likely to lose.

Theoretical Models

17. Professional Arbitrage (Limits of Arbitrage II)

Reference:

Emmons, William R., and Frank A. Schmid (2002) "Asset Mispricing, Arbitrage, and Volatility," *Federal Reserve Bank of St. Louis Review*, Vol. 84(6), 19-28, http://research/stlouisfed.org/publications/review.

Shleifer, Andrei (2000) *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford: Oxford University Press, Chap. 4.

"Markets can remain irrational longer than you remain solvent."

—John Maynard Keynes

The nature of the arbitrage business

Convergence traders try to exploit asset mispricing—that is, financial market inefficiencies—by going long on the comparatively underpriced security and short on the comparatively overpriced security

A hedged trading position is self-financing because the short-sale proceeds can be used to finance the long position

Short-selling earns the risk-free rate—as the proceeds of the short-sale are invested at the risk-free rate—plus the change in the price of the security

Interest payments or dividend payments of the borrowed security go to the securities lender.

The opportunity cost of the long leg equals the return of the proceeds from the short-sale—in other words, the two legs finance each other

If financial markets are efficient, the return on long-short portfolios (that is, hedge portfolios) is zero.

Today, most finance scholars agree that there is a modicum of inefficiency and predictability in financial markets, originating from temporary asset mispricing that inevitable and predictably disappears over time

Convergence trading tries to profit from such asset mispricing

Unfortunately, convergence trading is no free lunch

First, there are transaction costs

Second, there is the risk that asset mispricing might deepen temporarily, which poses a liquidity risk

Third, there is a risk that the convergence trader errs

There should always be a modicum of doubt about one's own beliefs, acknowledging Knightian uncertainty.

Fourth, there is a problem of asymmetric information between the hedge fund manager and investors.

The business of convergence trading is often times organized into hedge funds

- The term "hedge fund" is a misnomer because the business of hedge funds is arbitrage, not hedging
    - Note that hedge portfolios are hedged portfolios
        - For instance, an investor who goes long on a stock of a given CAPM beta and goes short, at the same amount, on another stock of an identical CAPM beta, holds a <u>market-neutral</u> portfolio.
- Hedge funds are private, and thus unregulated
    - Typically, an investors has to contribute a minimum of $1 million
        - Some hedge funds have lock-up periods of one to three years, whereas others allow investors to withdraw money with only a few weeks' notice.
    - Hedge funds typically charge their investors a fixed annual "management fee" of one or two percent of assets under management plus an "incentive fee" of 15 to 25 percent of the fund's realized annual return
        - The incentive fee is waived if a particular "hurdle rate" has not been achieved, which can be a fixed number or a reference rate such as the Treasury-bill rate plus or minus a spread
        - Most funds also apply a "high-water mark" provision, which requires the fund to make up any past losses before the incentive fee is paid.

Hedge funds show a very wide dispersion of returns in cross-section (and over time)

- The performance of the hedge fund industry is difficult to ascertain because of the survivorship bias
  - End-of year performance statistics tends to ignore hedge funds that went out of business during the calendar year in question.
  - Performance evaluations that try to account for the survivorship bias find that hedge fund investors should not expect more than the risk-free rate of return (after transaction costs and management fees)—empirical evidence that speaks against the hypothesis of market inefficiencies offering a free lunch.

A model of convergence trading

There are three types of agents in the model—all types of agents being risk-neutral

Noise traders have wealth but misperceive the intrinsic value of a given financial asset

Noise traders are unsophisticated investors—investors who are unable to separate signal from noise (and, consequently, trade frequently)

If noise traders beliefs are correlated in cross-section, their erroneous beliefs are sufficiently powerful to cause deviations of the market from intrinsic value—the Nasdaq and *Neuer Markt* being cases in point.

Professional arbitrageurs have no wealth or borrowing capacity but know the intrinsic (or "fundamental") value of the financial asset

Investors have wealth but no insight into the financial asset's intrinsic value

Unlike noise traders, investors *know* that they cannot recognize the asset's intrinsic value.

The arbitrageurs must convince wealthy but uninformed investors to entrust them with investment capital

Unfortunately, arbitrageurs cannot *prove* in advance that they recognize the intrinsic values of the assets they claim are mispriced

Even worse, it is possible that the assets will become even more mispriced before reverting eventually to intrinsic value

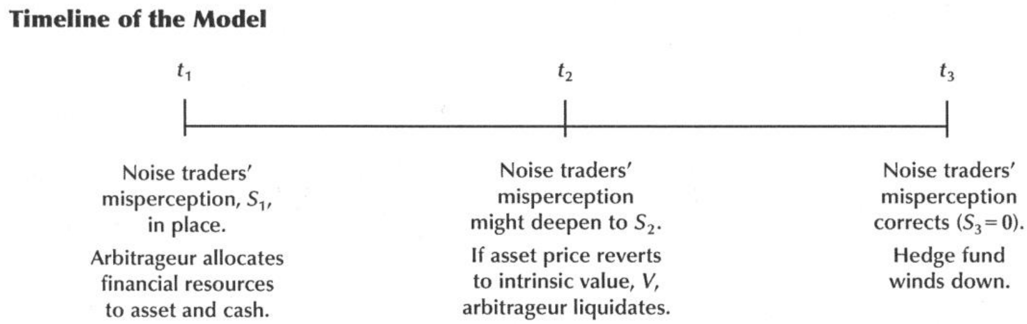
In such an event, after having suffered losses, the outside investors may demand their money back even though the expected profit of staying invested <u>has actually increased</u>.

The model shows that the existence of professional arbitrageurs mitigates—but cannot eliminate—mispricing in the market, regardless of how sensitive the outside investors are to arbitrageurs' past performance in deciding whether to remain invested with them after suffering losses

The model also shows that arbitrage dampens the <u>unconditional volatility of asset returns</u>—measured as the expected value of squared returns

Critical to the results is the arbitrageurs over time <u>accumulate wealth</u>—wealth that they invest in the hedge funds because they know the intrinsic value of their portfolios.

Timeline



The asset trades at three moments in time, t $(t = 1,2,3)$

We capture the influence of the noise traders' misperceptions of the intrinsic value of the asset at times t_1 and t_2 with the parameters S_1 and S_2, respectively

There is no fundamental risk in the model because the price of the asset will revert to intrinsic value at a known date (t_3) with certainty (so $S_3 = 0$).

Although the investment is risk-free from a two-period perspective, it is risky from a one-period perspective because the asset mispricing might deepen temporarily

In response to a deepening of the mispricing and the ensuing loss of the hedge fund the outside investors might revise their beliefs about the arbitrageur's abilities and withdraw money precisely when the expected return on the arbitrage is at its maximum, terminating their investment with a loss.

The underpriced asset in question is a long-short (that is, hedge) portfolio

> Note that going long on a comparatively undervalued asset and going short on a comparatively overvalued asset means holding an undervalued asset (portfolio).

We assume that arbitrageurs are highly skilled people who pursue proprietary trading strategies and therefore enjoy a monopoly in their segments

> The hedge fund moves the asset prices in the respective market segment or, in other words, operates in an imperfectly liquid market.

For simplicity only, we make the assumption that the operating costs in the arbitrage industry are zero

Also, for simplicity, we set the risk-free rate of return and the marginal opportunity cost of capital equal to zero

The supply of the financial asset is unity

Noise traders' demand for the financial asset at time t $(t = 1,2,3)$ is expressed as

$$QN_t = \frac{V - S_t}{p_t}$$

- where $0 < S_t < V$ for $t = 1,2$ and $S_3 = 0$, p_t is the price of the financial asset, and S_t is the misperception of the noise traders about the intrinsic value of the financial asset
- Remember that the asset is a long-short portfolio, which must be underpriced if the noise traders misperceive the intrinsic value of at least one leg of this portfolio
- Without misperception ($S = 0$), the noise traders would be willing to absorb the unit supply of the asset or, in other words, the asset would trade at intrinsic value ($p_t = V$).

The arbitrageur is compensated in two ways in accord with actual practice—via an up-front "management fee" and an after-the-fact performance-based "incentive fee"

- At the beginning of each period, he receives a fraction (α) of the assets under management, and at the end of the period he receives the fraction (β) of any positive return on the portfolio
  - This corresponds to compensation structures in real-world hedge funds, where managers typically collect $\alpha = 1$ percent or $\alpha = 2$ percent of the equity capital, plus $\beta = 20$ percent of any positive return on the fund's equity.
- We assume that the arbitrageur invests his entire fee income in the fund.
  - This is because the arbitrageur recognizes the profitability of the fund's activities.

The variable F_t denotes the total financial resources available to the arbitrageur at time t $(t = 1,2,3)$

- The value of F_1 is exogenous, whereas the quantities F_2 and F_3 are determined in the model
  - The startup capital, F_1, is provided solely by the investors, whereas the arbitrageur acquires the share α in F_1 immediately as part of his compensation.
- The arbitrageur acquires additional equity at t_2 in the amount of a fraction α of the outsiders' share in F_2
  - Furthermore, the arbitrageur acquires equity in the fund through capital gains on his equity position and through his share β in the capital gains on the outsiders' equity.
- The quantity F_3 is the fund's liquidation value.

We assume that the fund raises equity capital only at the outset (that is, at t_1)—an assumption we motivate as follows

- The arbitrageur's compensation depends not only on the return on his own equity stake but also on the amount of (and return on) the outsiders' equity capital under management
  - The arbitrageur therefore might have an incentive to raise fresh capital at t_2, particularly if he expected low returns in the second period
    - The injection of fresh funds would dilute the fund's existing investors' equity stakes.
- Thus, we assume—in keeping with typical hedge-fund arrangements—that the fund closes to <u>new and existing</u> investors after raising the initial capital.

At time t_2, the price of the asset either reverts to V or it does not

If the asset price is V at t_2, the arbitrageur liquidates the fund and holds cash until t_3

If the asset price does not equal V at t_2, the arbitrageur invests aggressively—albeit not all of the fund's cash—in the underpriced asset

This second-period investment generates a risk-free return because the asset price will rise to V at t_3 for sure.

The arbitrageur's (that is, hedge fund's) demand for the asset at the interim date, t_2, is given by

$$QA_2 = \frac{D_2}{p_2} \ , \ 0 \le D_2 \le F_2 \ ,$$

where D_2 is the amount of the hedge fund's demand in dollars; the amount $F_2 - D_2 \ge 0$ is held in cash.

Because total demand aggregated across noise traders and the arbitrageur must equal the asset supply of one unit ($QN_2 + QA_2 = 1$), the price of the financial asset at t_2 is determined by

$$p_2 = V - S_2 + D_2 \ , \ 0 \le D_2 < S_2 \ .$$

The condition $D_2 < S_2$ implies that the asset still trades at a discount to intrinsic value at t_2: $p_2 < V$

This assumption recognizes the arbitrageur's incentive *not* to bid up the price all the way to intrinsic value immediately

The arbitrageur would forego a positive return on investment in the second period.

With D_1 denoting the amount the arbitrageur invests in the asset at t_1, we have

$$QA_1 = \frac{D_1}{p_1} \ ,$$

which implies the initial asset price will be

$$p_1 = V - S_1 + D_1 \ , \ D_1 < S_1 \ .$$

The condition $D_1 < S_1$ implies $p_1 < V$, which again captures the fact that the arbitrageur will not bid the price all the way up to the asset's intrinsic value due to the incentives built into his compensation schedule

After all, the asset may become even more underpriced at t_2, in which case he will want to increase his investment ("double up").

The investors have prior beliefs about the arbitrageur's talent in exploiting possible asset mispricing, but are not perfectly informed

Investors update their beliefs about the arbitrageur's talent using a simple Bayesian learning rule, which is based solely on the arbitrageur's past performance

When past returns are poor, investors don't know for sure whether the poor returns are due to a random error (noise), a deepening of noise-trader misperception (bad luck), or truly inferior investment talent—a problem of observational equivalence

Pulling some of their money from the hedge fund after the asset mispricing has deepened—that is, when the expected return on the long-short portfolio is highest—is the investors' rational response to the loss.

The investor's rule of updating his beliefs about the arbitrageur's talent implies that, if the hedge fund loses money during the first period, the fund faces withdrawals at the interim date, t_2

Specifically, we assume that the withdrawals at t_2 are a multiple of the hedge fund's posted gross return (that is, before management fees) at t_2, denoted R_2, should this return be negative

Thus, the supply of funds in the second period is the following:

$$F_2 = \begin{cases} F_1 \cdot \alpha \cdot (1+R_2) + F_1 \cdot (1-\alpha) \cdot (1+R_2)^{\gamma} \;\;, \gamma > 1, & \text{if } -1 \leq R_2 < 0 \\ F_1 \cdot (1+R_2), & \text{if } R_2 > 0 \end{cases},$$

where γ is a parameter that determines the responsiveness of the investor to past performance

For $\gamma = 1$, poor first-period returns do not shake the confidence of investors in the arbitrageur's talent

At the other extreme, responsiveness that becomes unboundedly large implies that even a small first-period loss is multiplied into a huge withdrawal of funds

Note that the outside investors may withdraw only what is theirs

This means that, even if the outsiders pull all of their money, the arbitrageur's equity stake remains and the fund can stay in business

The arbitrageur knows that—despite a temporary deepening of the mispricing—the price of the asset will revert to intrinsic value at t_3 for certain, so he will keep his own money invested, come what may.

The gross return of the hedge fund in the first period, R_2, is given by

$$R_2 = \frac{(F_1 - D_1) + D_1 \cdot \frac{p_2}{p_1} - F_1}{F_1} \ .$$

For simplicity, we assume a specific form of uncertainty about noise trader sentiment at t_2, S_2

With probability $1-q \quad (0<q<1)$, noise traders recognize the true value of the asset, which implies $S_2 = 0$

In this case, the arbitrageur liquidates at t_2 and holds cash until t_3

Then the arbitrageur's assets under management at t_3 would amount to

$$F_3^{S_2=0} = F_2^{S_2=0} \equiv F_1 \cdot (1 + R_2^{S_2=0}) \ ,$$

$$\text{where } R_2^{S_2=0} = \frac{(F_1 - D_1) + D_1 \cdot \frac{V}{p_1} - F_1}{F_1} \ .$$

On the other hand, noise trader misperception deepens to S_2 with probability q, $S_2 = S > S_1$ (> 0)

If noise traders continue to misperceive the intrinsic value of the asset, the hedge fund's assets at t_3 will amount to the following:

$$F_3^{S_2 = S} = \frac{V}{p_2^{S_2 = S}} \cdot D_2 + (F_2^{S_2 = S} - D_2)$$

$$= \frac{V}{p_2^{S_2 = S}} \cdot D_2 + F_1 \cdot (1 + R_2^{S_2 = S})^{\gamma} - D_2 \ ,$$

where $R_2^{S_2 = S} = \dfrac{(F_1 - D_1) + D_1 \cdot \dfrac{p_2^{S_2 = S}}{p_1} - F_1}{F_1}$.

The arbitrageur's total income consists of management fees and capital gains on reinvested management fees

The expected value of the management fees, MF, equals the sum of the expected values of the management fees collected at t_1, MF_1, at t_2, MF_2, and at t_3, MF_3; the expected value of the capital gains is CG

The arbitrageur's maximization problem therefore reads

$$\underset{D_1, D_2}{Max}\ \{MF_1 + MF_2 + MF_3 + CG\}\ ,$$

where management fees are

$$MF_1 = \alpha \cdot F_1\ ,$$

$$MF_2 = MF_2^{S_2=S} + MF_2^{S_2=0}, \text{ and}$$

$$MF_3 = \beta \cdot q \cdot R_3^{S_2=S} \cdot (F_2^{S_2=S} - (1+R_2^{S_2=S}) \cdot MF_1 - MF_2^{S_2=S})$$

and where

$$MF_2^{S_2=S} = \alpha \cdot q \cdot (F_2^{S_2=S} - \alpha \cdot [1+R_2^{S_2=S}] \cdot F_1 - \beta \cdot \max\{0, R_2^{S_2=S}\} \cdot (1-\alpha) \cdot F_1)$$

$$+ \beta \cdot q \cdot \max\{0, R_2^{S_2=S}\} \cdot (1-\alpha) \cdot F_1,$$

$$MF_2^{S_2=0} = \alpha \cdot (1-q) \cdot (F_2^{S_2=0} - \alpha \cdot [1+R_2^{S_2=0}] \cdot F_1 - \beta \cdot R_2^{S_2=0} \cdot (1-\alpha) \cdot F_1)$$

$$+ \beta \cdot (1-q) \cdot R_2^{S_2=0} \cdot (1-\alpha) \cdot F_1, \text{ and}$$

$$R_3^{S_2=S} = \frac{(F_2^{S_2=S} - D_2^{S_2=S}) + D_2^{S_2=S} \cdot \dfrac{V}{p_2^{S_2=S}} - F_2^{S_2=S}}{F_2^{S_2=S}}\ .$$

The quantity $MF_2^{S_2=S}$ represents the income the arbitrageur collects at t_2 should the noise traders' misperception deepen in the first period, and $MF_2^{S_2=0}$ is the fee income if the asset reverts to intrinsic value

The arbitrageur also captures capital gains on the equity he builds from the reinvested management fees. The expected value of the capital gains, CG, equals:

$$CG = q \cdot (R_2^{S_2=S} \cdot MF_1 \cdot [1 + R_3^{S_2=S}] + R_3^{S_2=S} \cdot MF_2^{S_2=S}) + (1-q) \cdot R_2^{S_2=0} \cdot MF_1 \ .$$

The arbitrageur's choice variables are D_1 ($\leq F_1$) and D_2 ($\leq F_2^{S_2=S}$), which are the amounts the arbitrageur invests in the asset at t_1 and t_2, respectively

Unless the asset reverts to intrinsic value at t_2 ($p_2 = V$), the t_2-price of the asset (see above) is a function of the t_2-choice variable, D_2

Similarly, the t_1-price of the asset (see above) is a function of the t_1-choice variable, D_1.

We solve the maximization problem numerically

We hold constant all of the following:

$$V = 1;\ F_1 = S_1 = 0.2;\ S_2 = 0.4;\ q = 1 - q = 0.5;\ \alpha = 0.02;\ \beta = 0.2$$

Note that $F_1 = S_1 = 0.2$ means that the arbitrageur has sufficient buying power to eliminate the t_1-mispricing entirely if so desired

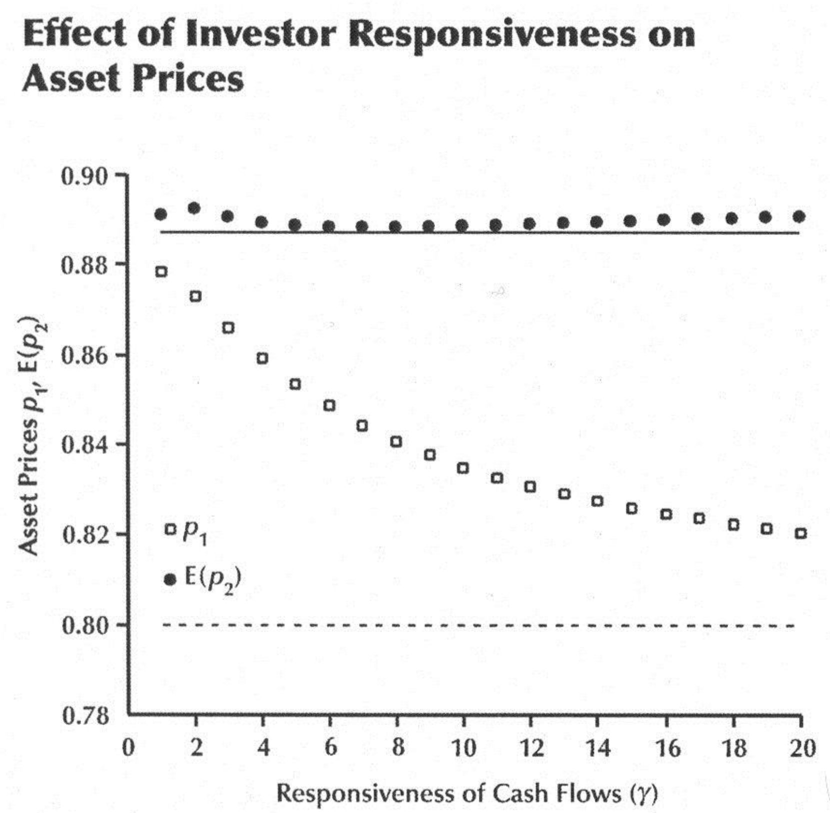
Also, note that $0.4 = S_2 > S_1 = 0.2$ means that noise trader misperception may deepen between t_1 to t_2—that is, the asset may become even more mispriced

For the values chosen for S_1, S_2, and q, noise trader misperception, S, is as likely to double as it is to vanish

Thus, the expected value of noise trader misperception in the second period, $q \cdot S_2$, equals the noise trader misperception observed in the first period, S_1.

We vary γ, the responsiveness to past performance of fund withdrawals, from $\gamma = 1$ (no responsiveness by the investors to past investment performance, that is, no withdrawals) to $\gamma = 20$ (extreme responsiveness) with a unit step length

The following chart shows the extent to which the hedge fund affects asset mispricing—the chart shows that the mispricing is less pronounced in each period than it would be without the hedge fund



Remember that, without arbitrage, the first-period price, p_1, and the expected value of the second-period price, E[p_2] both would equal 0.8 (shown as dashed line)

On the other hand, without noise traders, the asset would trade at unit value in both periods (not shown).

The hedge fund almost halves the difference between the expected value of the second-period price, $E[p_2]$ (shown as solid circles) and the asset's intrinsic, unit value

In fact, the degree of investor responsiveness, γ has little bearing on $E[p_2]$, which approaches the value of approximately 0.8873 (shown as a solid horizontal line) as γ approaches infinity

By comparison, the degree of responsiveness has a strong impact on the first-period price, p_1 (shown as open boxes)

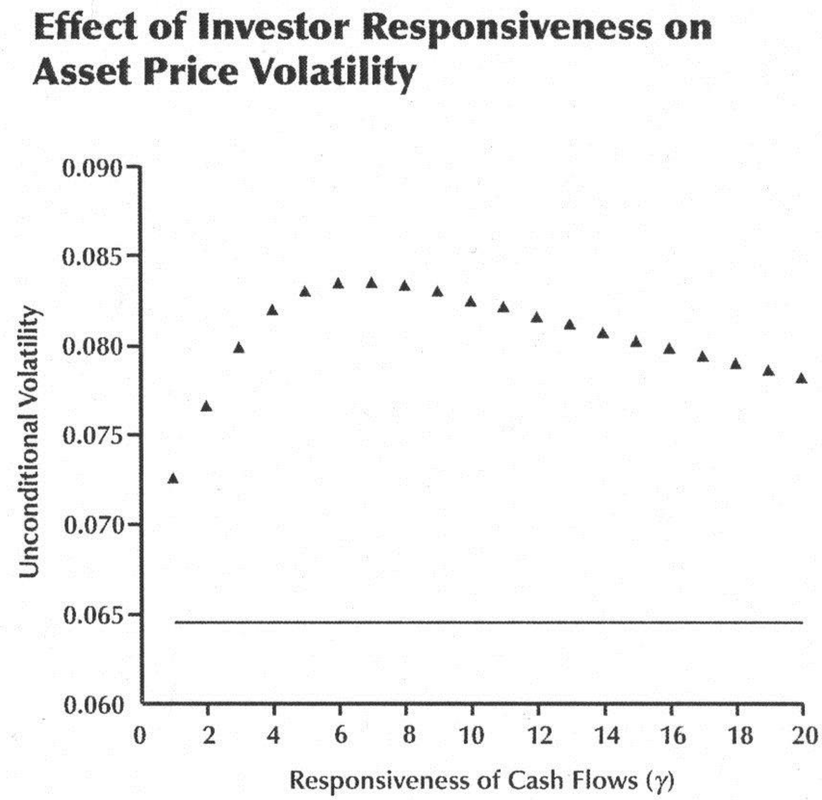
This is because the arbitrageur treads cautiously when putting on this trade in the first period when he knows that the investors penalize negative returns with sizeable withdrawals

In fact, the higher is γ, the more cash the arbitrageur holds in the first period, and therefore, the lower is p_1.

As the degree of investor responsiveness, γ, goes to infinity, the amount the arbitrageur invests in the first period goes to zero and consequently, the first-period price, p_1, converges to 0.8—the value the asset would adopt if there was no hedge fund in the market (shown as dashed line)

Thus we conclude that the hedge fund pushes the price of the asset (or its respective expected value) toward the intrinsic, unit value in both periods.

The following chart shows the unconditional volatility of the asset's returns for various degrees of investor responsiveness, γ



The unconditional volatility is calculated as the expected value of the squared returns over the two periods

For low values of investor responsiveness, volatility increases as γ increases

For high values of responsiveness, a further increase in γ reduces volatility monotonically

As γ goes to infinity, volatility approaches a level (as shown by the solid line) that is lower than the volatility level at $\gamma = 1$ (as indicated by the leftmost symbol), which is the benchmark case of unwavering investor confidence in the hedge fund manager

The reason for the "volatility hump" lies in the existence of two opposite effects

All else equal, the higher γ is, the bigger is the drop in the asset's price from t_1 to t_2, should the noise traders' misperception deepen

On the other hand, the higher γ is, the lower is the price of the asset at t_1 because the arbitrageur puts less money to work

For low values of investor responsiveness, the volatility-increasing effect dominates

For increasingly higher values of γ, this effect becomes progressively weaker until it vanishes for an infinitely large degree of investor responsiveness.

It is important to note that the hedge fund greatly reduces asset price volatility, regardless of the degree of investor responsiveness

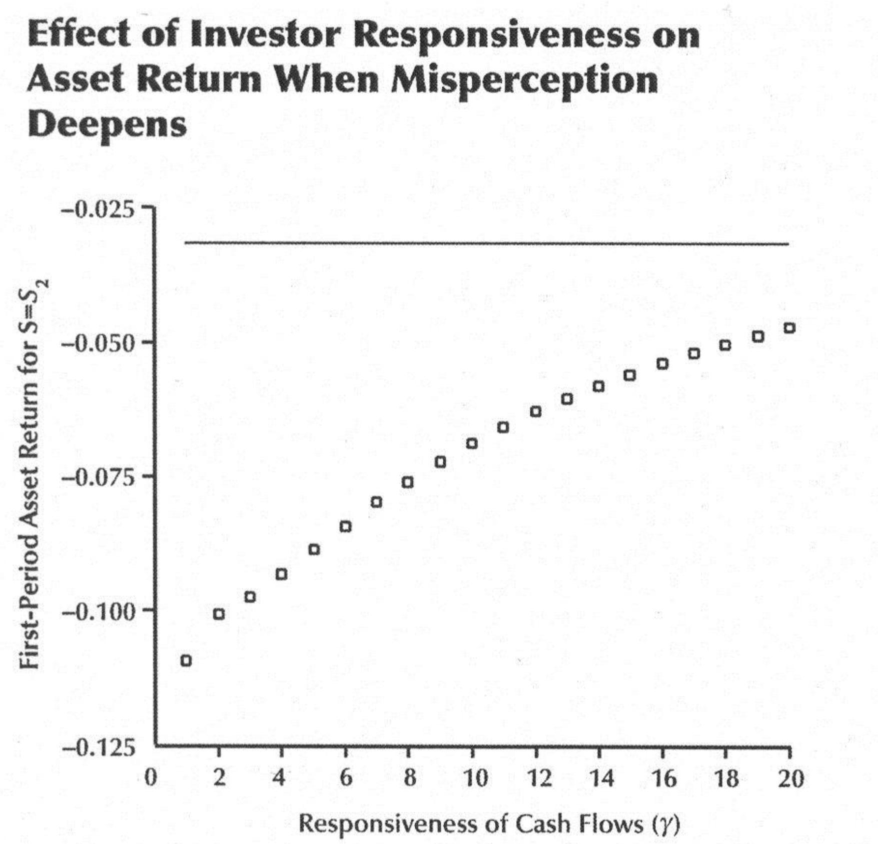
The unconditional volatility without the hedge fund runs at 0.5694 (not shown), which is a multiple of the volatility that we observe even at the degree of responsiveness that generates the highest level of volatility

Thus, we conclude that the hedge fund unambiguously reduces unconditional volatility.

Another way to look at the impact of arbitrage on volatility is ask how the market behaves in the event of a deepening of asset mispricing

Such an event—if severe—might cause, or occur alongside a financial crisis

The following chart shows for the case of a <u>deepening noise trader misperception</u> of the asset's intrinsic value the <u>second-period asset return</u> as a function of investor responsiveness



The absolute value of the percentage asset price decline increases with investor responsiveness, γ

For an infinitely high value of γ, the arbitrageur holds cash in the first period and then invests aggressively at t_2, although he does not invest all the cash available

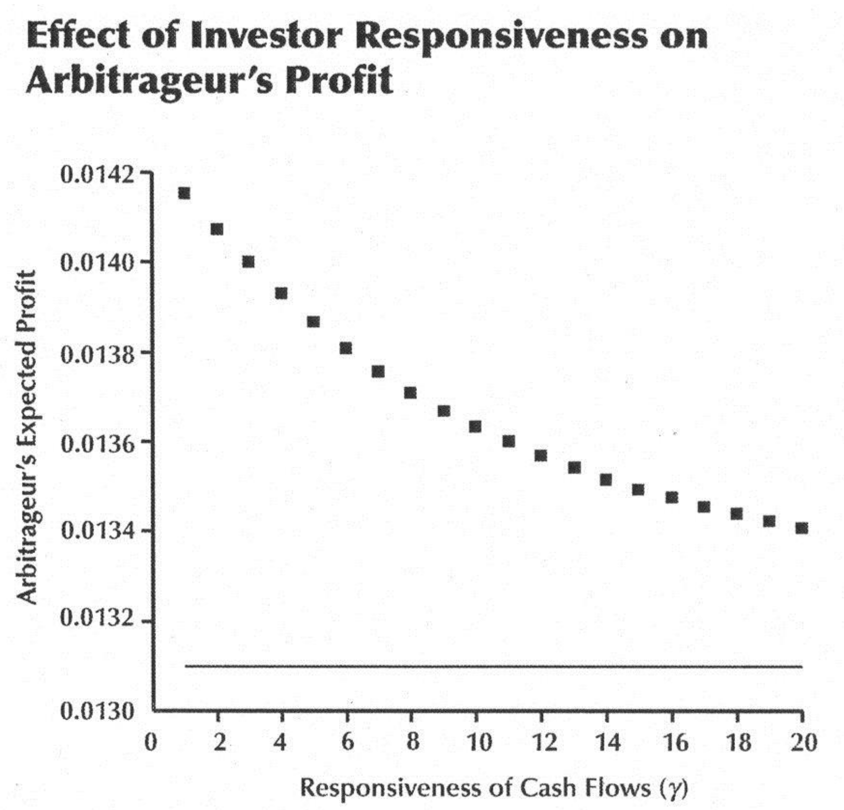
The horizontal line signifies the first-period return for this borderline case of an infinite degree of responsiveness

Note that, without a hedge fund, the first-period return would amount to minus 25 percent (not shown), which is more than twice as much (in absolute value) as what is observed with even a degree of responsiveness of zero

Hence we conclude that the presence of a hedge fund dampens volatility in the event of a deepening of noise trader misperception as might occur in a financial panic.

Finally, we are interested in the question of how investor responsiveness affects the arbitrageur's profit, that is, his incentive to set up a hedge fund and engage in arbitrage

The following chart shows the arbitrageur's profit as a function of γ



Not surprisingly, the profit of the arbitrageur decreases monotonically with increased investor responsiveness to past performance

Profit approaches a positive limit from above as the degree of responsiveness goes to infinity

The monotonic decline in the profitability of arbitrage with increasing investor responsiveness to past performance is a manifestation of the fact that liquidating a hedge portfolio when the expected return from arbitrage is highest is counterproductive—that is, it runs against "the nature of the trade."

[Blank page]

Theoretical Models

18. Feedback Strategies (Bubbles)

Reference:

Shleifer, Andrei (2000) *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford: Oxford University Press, Ch. 6.

Previously we looked at underreaction to earnings announcements as a source of momentum in stock prices (which generates predictability in stock returns)

Another possible cause of momentum in stock prices is feedback trading

In feedback trading, investors initiate trades based on past asset returns (rather than signals about the assets' fundamentals).

Possible, not mutually exclusive causes for feedback trading

Extrapolative expectations about prices

Extrapolative expectations of noise traders might give rise to momentum investing ('opportunistic investing') by arbitrageurs.

Stop-loss orders (long or short positions) or margin calls

Portfolio insurance

Portfolio insurance was a popular investment strategy prior to the 1987 stock market crash whereby institutional investors mimicked protective puts by continuous selling into falling prices

Portfolio insurance might lead to cascading prices

Portfolio insurance not only exacerbated the 1987 stock market crash; it proved ineffective because investors tried to sell at a time when buyers were hard to find (due to a dearth of liquidity).

In the following, we look at the consequences of noise traders having extrapolative expectations, which give arbitrageurs an incentive to pursue momentum strategies (opportunistic investment strategies)

Noise traders

Remember that if the stock price follows a random walk, the sequence of the six consecutive daily returns $+,+,+,+,+,+$ is as likely as the sequence $+,-,+,-,+,-$

Noise traders, who do not know the fundamental value of the asset and might not understand probability theory, erroneously believe that the sequence of positive returns reflects an increase in the asset's intrinsic value

Noise traders (in this context) are feedback traders; they react to past price changes (only), rather than to news that pertain to the intrinsic value of the asset

If noise traders' erroneous beliefs are correlated, their behavior is observationally equivalent to herd behavior, i.e., to behavior that follows the principle 'there is truth in numbers'.

Arbitrageurs

Arbitrageurs know the asset's intrinsic value and anticipate the noise traders' feedback trading

Arbitrageurs buy the asset in anticipation of the noise traders' feedback rule, which might add to the momentum in the asset price

Remember that although the arbitrageurs might aggravate a stock market bubble, it is the noise traders' erroneous beliefs that cause it

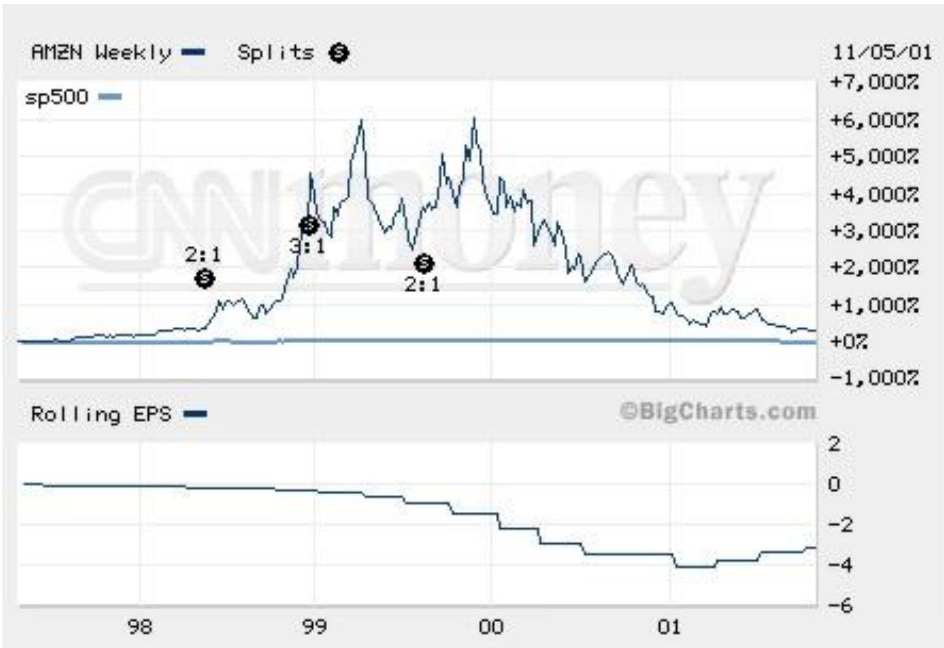
In the United States, stocks cannot be shorted on downticks; this measure is intended to prevent prices from cascading

(Downtick is the negative change in price between two subsequent trades; Nasdaq: change in bid price.)

An example of a bubble in the U.S. stock market is the dotcom bubble

The figure below shows the Amazon.com stock price (in percent of IPO price; weekly close; not total return) since going public on May 15, 1997

Note the S&P 500 price index at the bottom of the top panel, and the rolling (quarterly) earnings per share [$] data in the bottom panel; 'S' indicates stock splits



Source: http://money.cnn.com.

In early 1999, the (stock) market capitalization of Amazon.com reached $30 billion, increasing 20-fold since the beginning of 1998

Amazon.com lost about $90 million on selling $250 million worth of books per quarter over the Internet in 1998

The (stock) market capitalization of Amazon.com was seven times the combined market value of the two largest booksellers in the United States, *Barnes and Noble* and *Borders*, and 200 times their combined earnings.

Given the poor fundamentals of Amazon.com (whose earnings per share kept decreasing through early 2001), the run-up in the Amazon.com stock until March 2001 is hard to explain by underreaction to news about Amazon's fundamentals

A better explanation of the price bubble in dotcom stocks (as well as tech and telecom stocks) in the late 1990s is feedback trading

Positive-feedback investors buy securities after prices rise, and sell after prices fall.

History has seen many asset price bubbles, the most significant of which we will discuss at the end of this chapter

Walter Bagehot (1872, *Lombard Street*, London: Smith, Elder) reports

"… owners of savings … rush into anything that promises speciously, and when they find that these specious investments can be disposed of at a high profit, they rush into them more and more. The first taste is for high interest [dividends], but that taste soon becomes secondary. There is a second appetite for large gains to be made by selling the principal which is to yield the interest. So long as such sales can be effected the mania continues."

Arbitrage funds

In the chapters on noise trader risk (Limits of Arbitrage I) and on professional arbitrage (Limits of Arbitrage II) we took the position that arbitrageurs lean against noise trader demand and therefore stabilize prices, though sometimes imperfectly

(The only exception was a situation in the model on professional arbitrage where arbitrageurs exacerbated the price decline because they were forced to liquidate)

In the following model with feedback trading, however, arbitrage can be destabilizing (in the sense that arbitrage exacerbates the price effect of feedback trading).

George Soros

George Soros (1997, *The Alchemy of Finance*, New York: Simon and Schuster; 1998, *The Crisis of Global Capitalism*, New York: Public Affairs.) describes his own trading (investment) style as betting on future crowd [herd] behavior (as caused by correlated, possibly erroneous beliefs) rather than on fundamentals

Rather than selling short, Soros buys into bubbles early in the game and gets out before the bubbles deflates

Taleb (2001) reports that Soros—as a adherent to Popper's philosophy and unlike many other investors—retains a fair degree of skepticism about any investment style as he factors in the possibility of falsification

For instance, it has been argued that one of the reasons for the failure of Long-Term Capital Management (LTCM) was that it accounted for risk captured by their highly sophisticated models, but ignored the risk that the model itself might be wrong.

Overview on the model

There are noise traders, arbitrageurs, and 'passive' investors

Like arbitrageurs, passive investors have information about the intrinsic value of the asset at period 2, which allows them to buy low and sell high

Unlike arbitrageurs, however, passive investors do not receive a signal ε about the intrinsic value of the asset in period 1

When good news about the intrinsic value of the asset arrives …

… passive investors increase the demand for the asset because of improved fundamentals (a higher intrinsic value), which causes the price of the asset to rise

… arbitrageurs increase their demand in anticipation of increased demand by feedback traders in response to a price increase in the previous period

After observing the price increase that follows the good news about the asset's intrinsic value, the noise traders increase their demand for the asset

In anticipation of the noise traders' feedback trading, arbitrageurs drive up the asset price beyond intrinsic value

The increased demand by noise traders keeps the asset price above fundamental value even as the arbitrageurs sell out to the noise traders.

As the price of the asset returns to intrinsic value eventually, the noise traders sell out

While arbitrageurs (and passive investors) buy low and sell high, noise traders buy high and sell low.

The model in detail

Timeline

Period	Event	Positive Feedback Traders	Passive Investors	Arbitrageurs
		Total Demands of:		
0	None, benchmark period	0	0	optimally chosen (=0)
1	Arbitrageurs receive a signal ε of the period 2 fundamental shock Φ	0	$-\alpha p_1$	optimally chosen ($=D_1^a$)
2	Passive investors learn Φ	$\beta(p_1 - p_0)$	$-\alpha(p_2 - \Phi)$	optimally chosen ($=D_2^a$)
3	Liquidation: dividend Φ + θ declared, where θ is an unpredictable period 3 fundamental shock	$\beta(p_2 - p_1)$	$-\alpha(p_3-(\Phi + \theta))$	optimally chosen sets $p_3 = \Phi + \theta$

Quantities demanded by different classes of investors by period and events that reveal information to different classes of investors. β and α are parameters that determine the slopes of positive feedback traders' and passive investors' demand curves. p_0, p_1, p_2, and p_3 are asset prices in periods 0, 1, 2, and 3, respectively D_1^a and D_2^a are arbitrageurs' period 1 and 2 demands, respectively.

Source: Shleifer (2000, p. 159).

There are four periods (0,1,2,3) and two assets: cash and stock

Cash is in perfectly elastic supply and offers a zero return

Stock is in zero supply and might be thought of as bets different types of investors make against one another

The investors consume (and die) in period 3, which means that the stock is liquidated in period 3

In period 3, the stock pays a risky dividend equal to $\Phi + \theta$

The value of Φ becomes publicly known in period 2, and a signal ε about Φ is received by the arbitrageurs (and only the arbitrageurs) in period 1

The random variable Φ takes on the following three values with equal probability: $-\phi, 0, \phi$ (i.e., $E(\Phi) = 0$)

The value of θ becomes (publicly) known in period 3, and is normally distributed: $\theta \sim N(0; \sigma_\theta^2)$

There are three types of investors

Positive-feedback traders (which are noise traders), denoted f

Passive (informed) investors, whose demand for the asset in all periods depends only on the price relative to its fundamental value

Informed investors, denoted i, are present in a measure of μ

Arbitrageurs who maximize expected utility as a function of period 3 consumption

Arbitrageurs, denoted a, are present in a measure of $1-\mu$

The total of passive investors and arbitrageurs is held constant to derive comparative static results on the effect of changes in the number of arbitrageurs, holding constant the risk-bearing capacity of the market.

Period 3

There is no trading

Investors pay each other according to the positions they hold in the stock and the publicly known dividend $\Phi+\theta$

Because the dividend is known for certain in period 3, arbitrageurs pin down the stock price to its fundamental value of $\Phi+\theta$

Period 2

The value of Φ is revealed to the arbitrageurs and passive investors

Positive-feedback traders' demand for stock in period 2 is given by

$$D_2^f = \beta \cdot (p_1 - p_0) = \beta \cdot p_1$$

where p_1 is the price in period 1, p_0 is the price in period 0 (which is set equal to 0), and $\beta > 0$ is the feedback coefficient

Feedback traders place a market order today in response to a past price change

One way to describe this behavior is that investors react to past history of capital gains by raising their estimate of the mean rate of return and thus increase their demand.

Note that no rational investor would hold a positive quantity of stocks in period 2 if $p_2 > \Phi$, because such a portfolio has a negative expected return

In contrast, positive-feedback traders' demand is invariant to the period 2 price.

Arbitrageurs choose their period 2 demand, D_2^a, to maximize a <u>mean-variance utility function</u>:

$$\underset{D_2^a}{Max}\ \{E[D_2^a \cdot (\Phi + \theta - p_2)] - \gamma \cdot Var[D_2^a \cdot (\Phi + \theta - p_2)]$$

where $\gamma := \frac{r_A}{2} (> 0)$ is the Arrow-Pratt measure of <u>absolute risk aversion</u>

Note that maximizing a mean-variance utility function is equivalent to maximizing the <u>certainty equivalent</u> of wealth, which can be written as the difference between expected value and risk premium

Remember that mean-variance analysis is appropriate when returns are normally distributed (as they are in the model in question, by assumption) or investors' preferences are quadratic

H. Levy and H.M. Markovitz (1979, "Approximating Expected Utility by a Function of Mean and Variance," *American Economic Review* 69: 308-17) show that mean-variance analysis can be regarded as a second-order Taylor-series approximation of standard expected utility functions

The objective function is the <u>certainty equivalent of wealth</u> (which is equal to the difference between the expected value and the risk premium).

The optimal demand, which results from the first-order condition of the maximization problem above, reads:

$$D_2^a = \frac{\Phi - p_2}{2 \cdot \gamma \cdot \sigma_\theta^2} = \alpha \cdot (\Phi - p_2)$$

where $\alpha := 2 \cdot \gamma \cdot \sigma_\theta^2 (> 0)$ for notational convenience.

Passive investors, who differ from arbitrageurs in that they do not receive the signal ε in period 1, have the following period 2 demand:

$$D_2^i = \alpha \cdot (\Phi - p_2) = D_2^a$$

Remember that arbitrageurs are present in a measure μ, and passive investors are present in a measure $1-\mu$, which means that arbitrageurs' and passive investors' total demands equal $\mu \cdot D_2^a$ and $(1-\mu) \cdot D_2^i$, respectively.

Also, remember that changes in μ do not change the risk-bearing capacity of the market, which allows us to isolate the consequences of introducing arbitrageurs.

Stability condition

The model has stable solutions if (and only if) $\alpha > \beta$

The responsiveness of noise traders' demand to past price changes, β, must be smaller than the responsiveness of the demand the arbitrageurs' and passive investors' demand to current price changes, α.

Period 1

Arbitrageurs receive a signal, ε, about the expected fundamental value of the asset, Φ (which is the expected value of the dividend payment in period 3, $\Phi + \theta$)

The signal might be <u>noiseless</u>:

$\varepsilon = \phi$ implies $\Phi = \phi$ with certainty

… or <u>noisy</u>:

For $\varepsilon = \phi$, Φ either equals ϕ or 0, with equal probability, which implies $E[\Phi] = \frac{\phi}{2}$

For $\varepsilon = -\phi$, Φ either equals $-\phi$ or 0, with equal probability, which implies $E[\Phi] = -\frac{\phi}{2}$

Arbitrageurs choose their demand D_1^a to maximize the same mean-variance utility function as in period 2 over the distribution they face as of period 1 of their certainty-equivalent wealth in period 2

Passive investors' demand in period 1 is of the same functional form as in period 2 (which means that they buy low and sell high):

$$D_1^i = -\alpha \cdot p_1$$

Positive-feedback traders' demand in period 1 is equal to zero:

$$D_1^f = 0$$

Because feedback traders react to past, but not to current price change, positive-feedback traders do not trade in period 1.

Period 0

Period 0 is a reference period; no signals are received; the price is set to its initial fundamental value of zero (i.e., no one is betting against anyone else), and there is no trading

Because there is no trading, in periods 0 or 3, the market clearing conditions are automatically satisfied in those periods

For periods 1 and 2, the market clearing conditions read, respectively:

$$0 = D_1^f + \mu \cdot D_1^a + (1-\mu) \cdot D_1^i$$

$$0 = D_2^f + \mu \cdot D_2^a + (1-\mu) \cdot D_2^i$$

Solution with a noiseless (i.e., perfectly informative) signal

We consider a positive shock ($\varepsilon = \phi$, as mentioned):

The (period 1) signal $\varepsilon = \phi$—for it is noiseless—implies $\Phi = \phi$ (expected value of period 3 dividend payment) with certainty.

As long as arbitrageurs are present in positive measure ($\mu > 0$), their trades guarantee the equality of prices in periods 1 and 2

Arbitrageurs bid up the period 1 price all the way to what the feedback traders are willing to pay for the asset based on this very same price increase.

If no arbitrageurs are present ($\mu = 0$), then the period 1 price equals zero, for no one has information about the expected period 3 value of $\Phi = \phi$

Thus, we can write:

$$p_1 = \begin{cases} p_2 & \text{if } \mu > 0 \\ 0 & \text{if } \mu = 0 \end{cases}$$

Imposing market clearing in period 2, we obtain after inserting the period 2 demand functions derived above:

$$0 = \beta \cdot p_1 + \mu \cdot \alpha \cdot (\phi - p_2) + (1 - \mu) \cdot \alpha \cdot (\phi - p_2)$$

$$= \beta \cdot p_1 + \alpha \cdot (\phi - p_2)$$

Consequently, we can write:

$$p_2 = \begin{cases} \dfrac{\alpha \cdot \phi}{\alpha - \beta} = p_1 & \text{if } \mu > 0 \\ \phi & \text{if } \mu = 0 \end{cases}$$

Because of $\dfrac{\alpha}{\alpha - \beta} > 1$, the price of the asset is strictly above fundamental value ($p_2 > \phi$) in all periods (periods 1 and 2) when arbitrageurs are present than when they are absent

Remember, arbitrageurs are the catalysts of the mispricing, rather than the cause

The case of the mispricing is the positive-feedback trading of the noise traders, which the arbitrageurs try to turn into profit.

The degree of mispricing is independent of the measure μ at which arbitrageurs are present, $\mu > 0$

This peculiarity arises from the fact that the period 1 signal about Φ, ε, is noiseless

For an arbitrageur, the round-trip trade or purchasing in period 1 and selling in period 2 carries no risk since no uncertainty is resolved in period 2

Even a small measure μ of arbitrageurs is consequently willing to put on an arbitrarily large position.

For the price path when the signal ε is noiseless see the figure below

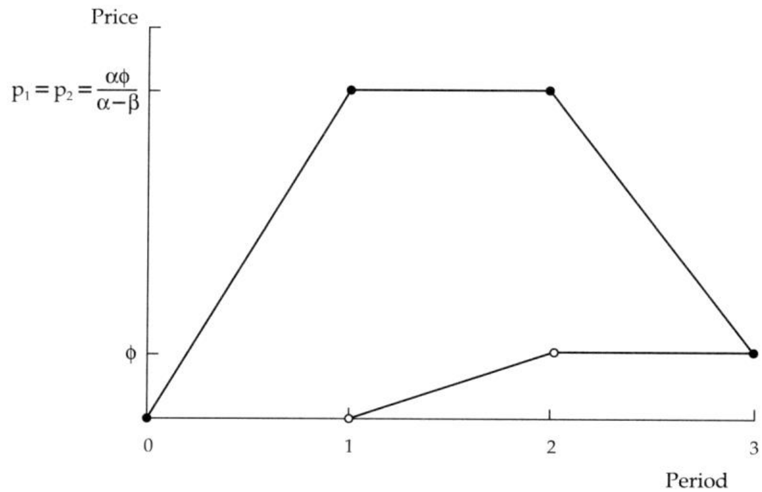


FIG. 6.1 Prices with a noiseless signal.
○ = Price without arbitrageurs. ● = Price with arbitrageurs. Pattern of prices with and without arbitrageurs in the noiseless signal model with a shock Φ to fundamentals. Arbitrageurs perceive the shock in period 1; passive investors perceive the shock in period 2 and fail to learn from period 1 prices. α is the slope of the demand curve of all non-positive feedback investors. β is the responsiveness of positive feedback investors' demand to past price changes.

Source: Shleifer (2000, p. 164).

Solution with a noisy (i.e., imperfectly informative) signal

We again consider a (possible) positive shock to fundamentals:

$$\Phi = \begin{cases} \phi & \text{with prob. } 1/2 \quad \text{[state A]} \\ 0 & \text{with prob. } 1/2 \quad \text{[state B]} \end{cases}$$

The period 2 market clearing condition (conditioned on the state of nature), reads:

$$0 = \begin{cases} \beta \cdot p_1 + \alpha \cdot (\phi - p_{2A}) & \text{[state A]} \qquad (1) \\ \beta \cdot p_1 - \alpha \cdot p_{2B} & \text{[state B]} \qquad (2) \end{cases}$$

The period 1 market clearing condition is given by:

$$0 = \mu \cdot D_1^a + (1-\mu)\cdot D_1^i$$

$$= \mu \cdot D_1^a - (1-\mu)\cdot \alpha \cdot p_1 \qquad (3)$$

where arbitrageurs' first-period demand D_1^a is still to be determined.

Before we determine the arbitrageurs' first-period demand, D_1^a, we write down his <u>expected</u> final (i.e., period 3) wealth as of period 2 for the two possible period 2 states of nature

For any <u>given</u> first-period demand of an arbitrageur, D_1^a, the arbitrageur's <u>expected</u> final (i.e., period 3) wealth <u>as of period 2</u> in the two possible states of nature, A and B, reads:

$$W_{2A}^a = D_1^a \cdot (p_{2A} - p_1) + \frac{\alpha \cdot (p_{2A} - \phi)^2}{2}$$

$$= D_1^a \cdot \left(\phi + \frac{\beta - \alpha}{\alpha} \cdot p_1 \right) + \frac{\beta^2 \cdot p_1^2}{2\alpha}$$

with $p_{2A} = \frac{\beta}{\alpha} \cdot p_1 + \phi$ resulting from the period 2 market clearing conditions above;

$$W_{2B}^a = D_1^a \cdot (p_{2B} - p_1) + \frac{\alpha \cdot (p_{2B})^2}{2}$$

$$= D_1^a \cdot \left(\frac{\beta - \alpha}{\alpha} \cdot p_1 \right) + \frac{\beta^2 \cdot p_1^2}{2\alpha}$$

with $p_{2B} = \frac{\beta}{\alpha} \cdot p_1 \equiv p_{2A} - \phi$ resulting from the period 2 market clearing conditions above

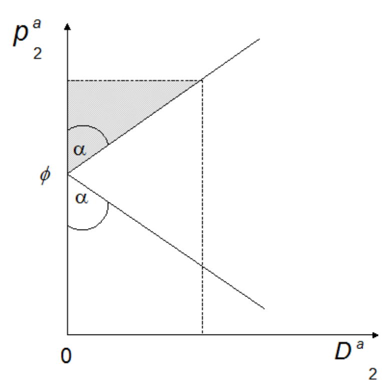
The <u>first terms</u> of the wealth functions, $D_1^a \cdot (p_{2A} - p_1)$ and $D_1^a \cdot (p_{2B} - p_1)$, respectively, represent an arbitrageur's profit from selling off to feedback traders in period 2 the position he acquired in period 1

- The arbitrageur's position appreciated from period 1 to period 2 due to the feedback traders' period 2 demand.

The <u>second terms</u> of the wealth functions, $\frac{\alpha \cdot (p_{2A} - \phi)^2}{2}$ and $\frac{\alpha \cdot (p_{2B})^2}{2}$, respectively, represent an arbitrageur's profit (<u>producer surplus</u>) from supplying the asset in period 2 to feedback traders in excess of the arbitrageur's existing position

- The arbitrageur <u>short-sells</u> the asset, i.e., promises to make the dividend payment in period 3, which amounts to $\Phi + \theta$, where Φ equals ϕ or 0, with equal probability

- The following figure depicts for state B the 'producer surplus' from short-selling (shaded area)



In period 1, the arbitrageur chooses the first-period demand D_1^a that maximizes the certainty equivalent (i.e., expected value minus risk premium) of his expected final wealth:

$$\underset{D_1^a}{Max}\ [\mathrm{E}(W_2) - \gamma \cdot \mathrm{Var}(W_2)]$$

where $\mathrm{E}(W_2) = D_1^a \cdot \left(\frac{\phi}{2} + \frac{\beta-\alpha}{\alpha} \cdot p_1\right) + \frac{\beta^2 \cdot p_1^2}{2\alpha}$ and

$$\mathrm{Var}(W_2) = (D_1^a \cdot \frac{\phi}{2})^2$$

The reason why maximizing the period 1 certainty equivalent of period 2 expected final wealth is methodologically unproblematic is that period 2 profit opportunities (short-selling) are independent of decisions made in period 1 or, in other words, period 3 risk is immaterial for positions taken in period 1

From the first-order condition of the maximization problem we obtain:

$$D_1^a = \frac{\frac{\phi}{2} + \frac{\beta-\alpha}{\alpha} \cdot p_1}{2\gamma \cdot (\frac{\phi}{2})^2}$$

$$= \frac{(p_{2A} + p_{2B}) - 2p_1}{\gamma \cdot (p_{2A} - p_{2B})^2} \qquad (4)$$

The arbitrageur's period 2 demand is simply the expected profit from holding one unit of the asset from period 1 to period 2, $\frac{\phi}{2} + \frac{\beta-\alpha}{\alpha} \cdot p_1$, divided by 2γ times the risk of holding this position, $(\frac{\phi}{2})^2$

Remember that 2γ is the arbitrageur's (absolute) degree of risk aversion.

Equations (1-4) form a system of four equations in four unknowns (the three prices p_1, p_{2A}, p_{2B}, and period 1 demand D_1^a); when solved for p_1, we obtain:

$$p_1 = \frac{\mu \cdot \phi}{2\mu \cdot \frac{\alpha - \beta}{\alpha} + \alpha \cdot \gamma \cdot \phi^2 \cdot (1-\mu)}$$

When there are no passive investors ($\mu = 1$), the price p_1 reads:

$$p_1 = \frac{\phi}{2} \cdot \frac{\alpha}{\alpha - \beta}$$

Without passive investors, arbitrageurs' period 1 holdings are zero, since there is no one from whom they can buy

For period 1 holdings to be zero, there must be no expected profit opportunity from buying in period 1 and selling in period 2. Hence, the period 1 price is simply the expected period 2 price.

When there are no arbitrageurs ($\mu = 0$), the price p_1 equals:

$$p_1 = 0$$

Without arbitrageurs, no one foresees the period 2 shock to fundamentals. Hence, the period 1 price is zero.

Remember that

$$p_{2A} = \frac{\beta}{\alpha} \cdot p_1 + \phi$$

$$p_{2B} = \frac{\beta}{\alpha} \cdot p_1$$

For $\beta > 0$, i.e., when there is positive-feedback trading, the period 2 price, p_{2A} or p_{2B}, always exceeds the period 1 price, p_1.

The figure below shows the effect of arbitrage on the price path when the signal is noisy

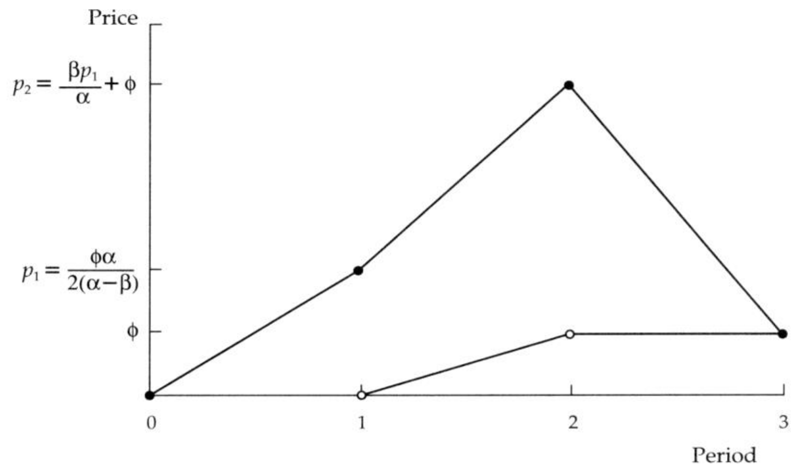


FIG. 6.2. Prices with a noisy signal.
○ = Price without arbitrageurs. ● = Price with arbitrageurs. Pattern of prices with and without arbitrageurs in the noisy signal model with a shock Φ to fundamentals. Arbitrageurs receive a noisy signal of the shock to fundamentals in period 1; passive investors perceive the shock in period 2 and fail to learn from period 1 prices. α is the slope of the demand curve of all non-positive feedback investors. β is the responsiveness of positive-feedback investors' demand to past price changes.

Source: Shleifer (2000, p. 167).

In period 1, arbitrageurs bet on Φ being high in period 2 ($\Phi = \phi$) and drive the period 1 price up above zero

This in turn raises positive-feedback trader demand in period 2 in both states of nature

The bet on future positive-feedback-trader demand drives the price in period 1 above its period 1 fundamental value of $\frac{\phi}{2}$

In period 2, arbitrageurs unload their positions and sell the asset short as positive-feedback demand keeps its price above the fundamental value

In summary, arbitrageurs buy in period 1, sell and go short in period 2, and cover (their shorts) in period 3

The model shows how arbitrage can destabilize prices

For $\mu > 0$ (i.e., when there are arbitrageurs in the market), the period 2 price is always further away from fundamentals than when $\mu = 0$, for, in the latter case, the asset trades at the (period 2) intrinsic value ϕ

Remember that without noise traders who engage in feedback trading, arbitrage would have no destabilizing impact.

However, if (and only if) $\frac{1-2\frac{\beta}{\alpha}}{\alpha \cdot \gamma \cdot \phi^2} < \frac{1-\mu}{\mu}$, the first-period price is closer to the first-period intrinsic value, $\frac{\phi}{2}$, with arbitrageurs than without

For sufficiently low values of μ, i.e., for sufficiently small measures of arbitrageurs, the first-period price is closer to $\frac{\phi}{2}$ than without arbitrageurs (in which case the first-period price equals zero).

Stock market bubbles (manias)

A stock market bubble—as illustrated in the above model—is characterized by a run-up with positively correlated returns, followed by mean reversion

There is casual evidence for noise trading and smart money responses in bubbles, as illustrated by the model discussed above.

There is a controversy about whether there is something like a stock market bubble

For instance, the price bubble in the theoretical model may be viewed as perfectly rational

All investors use the full set of information available to them at the time (albeit not all three types of investors are equally well informed).

While some finance scholar (for instance Charles Kindleberger, 1978, *Manias, Panics, and Crashes*, New York: Basic Books) view bubbles as outburst of irrationality associated with market inefficiency, other scholars (for instance, Peter M. Garber, 2000, *Famous First Bubbles*, Cambridge (MA): MIT Press) regard them as perfectly rational market outcomes

The 'rational' approach to bubbles is appealing for its adherence to strict academic principles, but its explanations often border on tautology.

Famous bubbles in history

Table 6.2 Famous bubbles

Bubble	*Initial displacement*	*Smart-money response*	*Sustaining the bubble*	*Authoritative blessing*	*Crash*	*Political reaction*
Dutch Tulipmania (1630s)	Mosaic viruses produce interesting looking tulips; prosperity of Holland	Selective breeding of tulips; purchase by 'insiders' of broken tulips that can only reproduce slowly and asexually	Development of tulip speculation contracts, which can be signed before notaries; appearance of trading	??	1637	??
South Sea Bubble (1710–20)	Profits from conversion of government debt; supposed monopoly on trade with Spanish ruled parts of America	Insiders buy up debt in advance of the conversion scheme, then profit by presenting debt for full conversion	Development of coffee house network for speculation; new subscriptions	Government approval; royal involvement	1720	ex post facto punishing directors; restrictions on use of the corporate form
Mississippi Bubble (1717–20)	Rapidly growing trade with the New World; Law's success as a financial organizer	Law's plan to make money and acquire power by securitizing the French debt	Government support; large expansion of credit by Law's bank to support further purchases	Official government support. Duke of Orleans imprisons critics of Law — the president of the Parlement de Paris and others	1720	Fall of Law; end of efforts to reform French finances until 1787
British first railway boom (1845–6)	End of depression; excitement over the new means of transportation	Many new railroad projects	Ponzi schemes by George Hudson (i.e. use this railroad's capital to pay the last railroads dividends)	Parliamentary bills passed for every railroad suggesting government approval; close links between George Hudson and London Society	No crash, gradual decline	Reform of accounting standards; requirement that dividends be paid only out of earnings, not out of capital
U.S. 1873 railway boom and crash	End of the Civil War; settlement of the American west	Construction of government subsidized railroads	Additional railroad charters; expectation that subsidies would continue	Henry Varnum Poor and Charles Frances Adams	1873 — Bankruptcy of Jay Cooke & Co., beginning of mid 1870s depression	??
Argentine loans (1880s)	Strong demand on world markets for the staple products of Argentinian agriculture; large profits made by early investors	Investment flows from Britain to Argentina; expansion of railway network; construction of of social overhead capital	New issues on the London exchange; creation of joint-stock companies to speculate in Argentinian land	Foreign investors 'grossly misled ... by Argentinian president' Barings' express optimism that the situation might improve (hoping to avoid bankruptcy)	Baring Bros. bankruptcy November 1890	Coup d'etat in Argentina; laws discriminating against foreign investment
1920s Florida land boom	Great winter climate; closeness to centers of American population; prosperity of the 1920s	Building of railroads; development of Miami; land development projects	Subdivisions; creation of a network of real estate offices selling Florida land	William Jennings Bryan boosts Florida land; close connections between mayors and developers	1926	Fraud prosecutions
1920s U.S. Stock Market boom	Decade of fast growth in the 1920s; end of fears of post WWI deflation; rapid expansion of mass production	Expansion of supply of shares; creation of new closed end funds	Regional exchanges; growth of margin accounts and brokers' loans	Blessings from Coolidge, Hoover, Mellon and Irving Fisher	October 1929 and following	Glass–Steagall Act; creation of SEC; public utility holding company act; election of FDR
1920s U.S. utility stocks boom	Expansion of demand for power; economies of scale	High leverage; expansion of scale to capture economies	Creation of public utility holding companies with cascades of control	??	October 1929 and following	Breakup of large utilities, TVA a byproduct; substantial government regulation of utility industries
1960s conglomerate mergers in the US	Two decades of a rising stock market during which investing in growth stocks had been profitable	Emergence of professional conglomerates; Harold Geneen's ITT, Textron, Teledyne, etc.	Stock swaps to create apparent earnings growth	Harvard endowment takes large positions in National Student Marketing; McGeorge Bundy urges institutions to invest aggresively	1970-1971	Reform of accounting practices; Williams Act

Source: Shleifer (2000, pp. 170-71).

Characteristics of bubbles

- Most price bubbles start with initial good news, which generate substantial profits for some investors in an asset
  - In the case of the Dutch Tulip (bulb) mania, mosaic viruses produced interesting looking tulips which connoisseurs paid high prices for
  - The Mississippi bubble was stimulated by expanding trade with the New World
  - The first British Railway boom arose out of excitement about the new means of transportation
  - The U.S. stock market boomed in the 1920s as the economy grew for several years at a remarkable pace.
- In response to the initial increases in asset prices, smart money (arbitrageurs) increases the supply of both the desirable physical assets and the claims to them
  - To increase the supply of actual assets that are the subject of excitement, smart money reproduces tulip bulbs, builds railroads, develops land, …
  - Smart money also creates financial assets to take advantage of noise traders, particularly in the later stages of the bubble as smart money tries to sustain the noise trader enthusiasm
    - Remember the chapter ‘The Long-Run Performance of IPOs’.

Often times, authoritative blessing exacerbates the bubble

- For instance, the Mississippi bubble received official government support, and in fact, some of the critics of John Law—the person that run the coffee house network for speculation—were imprisoned
- The famous economist Irving Fisher justified the boom in U.S. equities in the 1920s when declaring in the summer of 1929 that "… stock prices have reached a new and higher plateau"
- Did Federal Reserve Chairman Alan Greenspan justify the (as of December 2001 ongoing) boom in U.S. equities with remarks in the July 1997 Monetary Policy Report to the Congress that led to what is now called the 'Fed model' of stock market valuation?
  - For details see the chapter 'Stock Market Valuation'.

Inevitably, bubbles burst

Bubbles may deflate suddenly (in crashes), or slowly, or both

Although the U.S. stock market had depreciated by 39.6 percent between September 4, 1929, and October 29, 1929 (the day on which it fell 11.5 percent), it took the market almost three years to bottom out

Source: <http://www.stockpickssystem.com/stock_market_history.htm>.

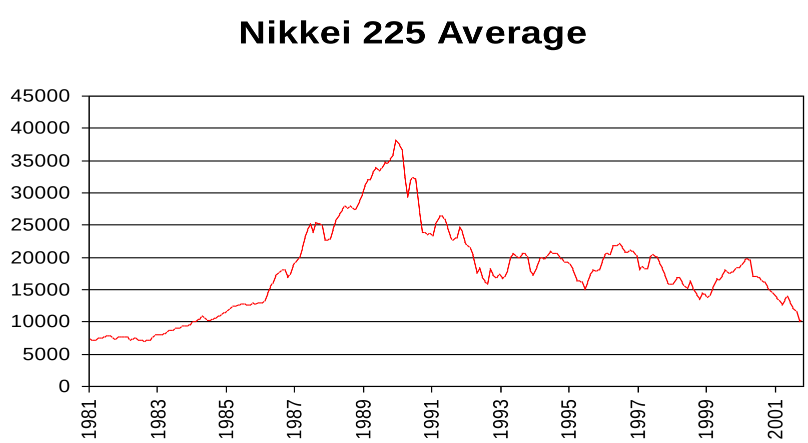
By July of 1932 the stock market would hit a low that made the 1929 crash look like hiccup

By the summer of 1932 the Dow had lost almost 89% of its value, which was well more than 50 percent lower than the low of October 29, 1929

This drop erased almost every gain from the stock market since its birth in 1897

It would take the U.S. stock market about 30 years to make it back to the 1929 highs, though most investors would have recovered their losses in the 1930's through dividends.

The Nikkei 225 took more than 11 years to descend from its December 29, 1989, peak of 38,915.87 (at close) to its (most recent) trough of 9,504.41 (at close) on September 17, 2001



Note: Monthly average; first observation: January 1981; last observation: October 2001; Source: Wall Street Journal.

Frequently, bubbles are followed by regulatory changes or even political instability

- The South Sea Bubble in England was followed by an over-century-long prohibition on the formation of joint stock companies without an explicit permission by Parliament
- The collapse of the 1880s Argentine loans bubble brought about a coup d'état, not to mention severe restrictions on foreign investment
- The stock market crash of 1929 resulted in the Glass-Steagall Act, the Public Utility Holding Company Act, the formation of the SEC (Securities and Exchange Commission, http://www.sec.gov) and many other regulatory changes of the financial markets.

Case Study

19. Valuation of 3Com and Palm

Reference:

Lamont, Owen A., and Richard H. Thaler (2001) "Can the market add and subtract? Mispricing in tech stock carve-outs," Working Paper, University of Chicago, http://gsb-www.uchicago.edu/fac/owen.lamont/research/wp.html.

Lamon and Thaler (2001) study stock carve-outs, which are cases in which a parent company floats part of the stock of a yet unlisted, wholly-owned subsidiary

- The authors discuss cases where such equity carve-outs lead to a violation of the law of one price
  - In an equity carve-out, holders of a share of company A (the parent company) are expected to receive x shares of company B (the subsidiary)
  - The authors study a 1998-2000 sample of carve-outs and discover cases of $P_{\mathrm{A}} < x \cdot P_{\mathrm{B}}$, where P_i is the share price of company i ($i = \mathrm{A,B}$)
    - A prominent example involves 3Com and Palm.
  - The authors find that arbitrage does not eliminate this blatant mispricing due to short sale constraints, i.e., transactions costs of short-selling B are prohibitive
  - Evidence from options pricing shows that costs of shorting are extremely high, thus eliminating arbitrage opportunities.

Efficient markets and the law of one price

In efficient markets, identical assets trade at the same price

The law of one price is expected to hold in financial markets where transactions costs are low and transparency is high

The driver of the law of one price is arbitrage, which—in its most basic manifestation—is defined as the simultaneous buying and selling of securitized claims on identical cash flow streams for two different prices (as a possibly self-financing portfolio)

Note that this (most basic) form of arbitrage has no fundamental risk as the cash flows of the securities are identical

Noise trader risk exists, that is, mispricing might deepen before the relative price of the two securities prices reverts to intrinsic value.

If efficient markets are defined as markets in which "deviations from the extreme version of the efficiency hypothesis are within information and trading costs" (Eugene F. Fama, 1991, "Efficient Capital Markets: II," *Journal of Finance* 46, 1575-1617), a market in which the law of one price is violated might still be efficient in the sense that there is no money left on the table

An example of an efficient market (albeit not in its extreme version) is one where a stock might be massively overpriced, yet, since there is no way for arbitrageurs to make money, the market might be efficient with grossly distorted prices.

3Com and Palm

Palm, which makes handheld computers, was owned by 3Com, a profitable company selling computer network systems and services

On March 2nd, 2000, 3Com sold a fraction of its stake in Palm to the general public via an initial public offering (IPO) for Palm shares

In this equity carve-out, 3Com retained ownership of 95 percent of the Palm shares

3Com announced that, pending an expected IRS (Internal Revenue Service) approval, it would eventually spin off its remaining shares of Palm to 3Com's shareholders before the end of the year

(Note that the spin-off is tax-free both to the parent firm and to its shareholders if the parent complies with Internal Revenue Code Section 355, which requires that the parent, prior to the spin-off, owns at least 80 percent of the subsidiary)

3Com shareholders would receive 1.525 shares of Palm for every share of 3Com they owned

Thus, there were two ways an investor could buy Palm:

The investor could buy, for instance, (roughly) 150 shares of Palm directly, or he could buy 100 shares of 3Com, thereby acquiring a claim to 150 shares of Palm plus a portion of 3Com's other assets

Because the price of 3Com's shares can never be less than zero (due to the put option embedded in stocks) the law of one price establishes the following inequality:

$$P_{3Com} > 1.5 \times P_{Palm}$$

Since 3Com at the time held more than $10 a share in cash and securities in addition to its other profitable business assets, one might expect 3Com's price to be well above 1.5 times the price of Palm.

Clearly, there is uncertainty about 3Com floating the remaining 95 percent of Palm

Yet, for those that consider Palm a hot commodity, the uncertainty about the remaining 95 percent of the stock is immaterial because they can acquire Palm assets simply by buying 3Com shares.

The day before the Palm IPO (which took place on February 29, 2000), 3Com closed at $104.13 per share

After the first day of trading, Palm closed at $95.06 per share, implying that the share price of 3Com should have jumped to at least $145 (using the precise ratio of 1.535)

- Instead, 3Com fell to $81.81, implying a 'stub value' of 3Com (the implied value of 3Com's non-Palm assets) of minus $23 per share
    - In other words, the stock market was saying that the value of 3Com's non-Palm business was minus $22 billion!
- Note that this stark mispricing is hard to explain by information costs, since the mispricing took place in a widely publicized IPO that attracted frenzied attention
    - The mispricing was not only simple in nature, it also was widely discussed, including in two articles in the *Wall Street Journal* and one article in the *New York Times*
    - Yet, the mispricing persisted for months!

Arbitrage opportunities

- The mispricing of 3Com and Palm relative to each other, created arbitrage opportunities
  - Remember that there is no fundamental risk (given the negative stub value of 3Com) and very little noise trader risk
    - As for noise trader risk, remember that when 3Com spins off the remaining 95 percent of shares, the relative mispricing goes away, that is, noise trader risk vanishes.
- An arbitrageur who buys 100 shares of 3Com and shorts 150 shares of Palm is essentially buying the 3Com stub for minus $63
  - At the time 3Com spins off the remaining 95 percent of Palm, 3Com's stub value must be at least zero
    - There is no fundamental risk because the value of Palm is perfectly hedged
    - Noise trader risk is of little importance, except if the arbitrageur faces the risk of having to close out due to unforeseen liquidity needs or because the broker loses the stock.
  - Yet, investors were willing to pay over $2.5 billion to buy the (relatively) expensive shares of Palm rather than the (relatively) cheap shares of Palm embedded in 3Com and get 3Com thrown in!
    - The authors show that the arbitrage opportunities could not be exploited (and the mispricing consequently could persist) because of high costs of shorting shares of Palm
    - Note that while short-sale constraints can explain why arbitrage worked so poorly, they cannot explain why investors wanted to buy the relatively expensive Palm shares (rather than the relatively inexpensive 3Com shares) in the first place.

Time path of stub values

The authors study a sample of 18 equity carve-outs, which are all those carve-outs from April 1996 to August 2000 where the parent company offered a definitive time schedule for distribution of the remaining shares

The authors calculate the stub value at any point in time t , S_t , between the carve-out and the distribution date (of the remaining shares):

$$S_t = P_t^P - x \cdot P_t^S, \text{ or in relative terms, } s_t = \frac{P_t^P - x \cdot P_t^S}{P_t^P}$$

where the superscripts P, S stand for the parent firm and the subsidiary, respectively

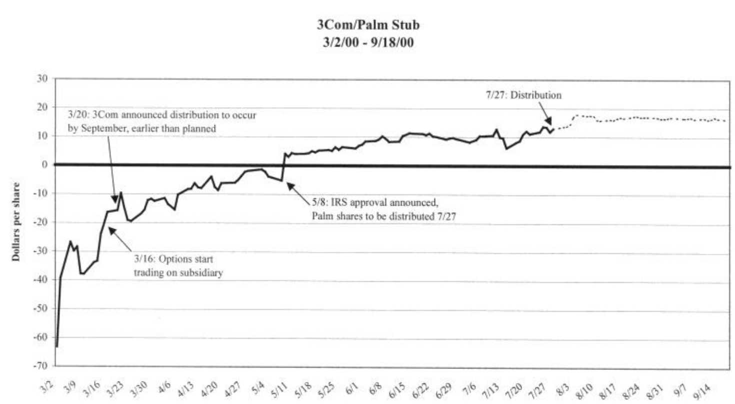
Calculating the stub value is plagued by uncertainty about the number of outstanding shares of the parent company, which might increase over time due to convertible debt or stock options

Thus the authors pursue only cases where they can unambiguously identify a negative stub; there are six such cases

In all but one case the uncertainty about the final ratio, x , caused by uncertainty about the number of outstanding shares of the parent company, appears to be small.

The authors' Figure 3 shows the time path of the 3Com stub, as an example

Figure 3



Risk and return in stub values

The authors study the risk and return characteristics of two investment strategies that go long on the parent and short on the subsidiary

The authors show that—on paper—the investment strategies produce high returns with relatively little risk

As mentioned, the strategies could—at the time—not be implemented because of high costs of short-selling, which in turn explains the high returns produced on paper.

In a first strategy ("simple strategy"), an investor shorts the subsidiary and uses the proceeds from the short-sale to finance the purchase of shares in the parent company

The portfolio is self-financing, and the exact distribution ratio of x is not important

The return on the portfolio during any time interval T equals:

$$R_T^P - R_T^S$$

where R^P, R^S are the returns on the stocks of the parent and the subsidiary, respectively.

In a second strategy, ("hedged strategy") an investor buys one share in the parent, shorts x shares of the subsidiary, and invests the net proceeds at the risk-free rate of return, R^F

For the hedge to be perfect, the distribution ratio of x must be known (which the authors assume in their calculations)

The portfolio is a pure bet on the stub

The return on the portfolio for any time interval T equals:

$$\frac{1}{1-s_0} R_T^P - \frac{s_0}{1-s_0} R_T^F - R_T^S$$

where $s_0 = \frac{P_0^P - x \cdot P_0^S}{P_0^P}$ is the stub value as a fraction of the parent stock price at the time the portfolio was initiated.

The estimation of a CAPM regression and a multi-factor model on either investment strategy for the six carve-outs that the authors identified as having a negative stub (at some point between the carve-out and the spin-off) are shown in the authors' Table 4

Table 4
CAPM and three factor regression for monthly trading strategies

	Simple strategy		Hedged strategy	
α	0.10	0.10	0.09	0.09
	(0.03)	(0.03)	(0.03)	(0.03)
RMRF	1.22	1.41	0.89	1.06
	(0.53)	(0.60)	(0.47)	(0.53)
HML		0.46		0.42
		(0.45)		(0.40)
SMB		0.47		0.43
		(0.63)		(0.56)
R^2	0.22	0.27	0.16	0.21

Monthly regressions of strategy returns on factors. Calculations use closing prices. The strategy takes a position on the last day of the month if the stub is negative on that day, and holds until the last day of the month prior to the distribution month. In all five cases, the position is initiated at the end of the first month of trading. Since Metamor/Xpedior does not have a negative stub at the end of the month, it is not included in this strategy. Equal weighted returns on from one to three paired positions per month. The simple strategy is $R_t^P - R_T^S$. The hedged strategy is $\frac{1}{1-s_0}R_T^P + \frac{-s_0}{1-s_0}R_T^F - R_T^S$. R_t^P is the monthly return from parent stock and R_T^S is monthly return from the subsidiary stock. $s_0 = \frac{P_0^P - xP_0^S}{P_0^P}$ is the stub value as a percent of parent stock price, as of the last day of the first month of trading. RMRF is CRSP value weighted market return minus Ibbotson T-bill return. HML and SMB are the value and size factors from Fama and French (1993) and come from the web page of Kenneth French. HML is the returns on stocks with high book to market ratios minus the returns on stocks with low book to market ratios. SMB is the return on small cap stocks minus the returns on big cap stocks. Number of observations is 21 months. Standard errors in parentheses.

Note the statistical significance of the intercept, α, which is called Jensen's alpha

Jensen's alpha reports an abnormal return (i.e., return above and beyond what corresponds to an efficient market given the risk-return profile of the portfolio) of 10 percent (simple strategy) and 9 percent (hedge strategy), respectively

In other words, the portfolios are located above the securities market line, which is at odds with the efficient markets paradigm.

- A caveat might be in order
  - The abnormal returns of the two arbitrage strategies might be an artifact of disregarding risk that is not captured by the CAPM or a multi-factor model
    - There is the risk that the IRS will disapprove a tax-free spin-off, in which event the parent company might cancel
    - The parent company might cancel the spin-off for other reasons, for instance because of a poor performance of the subsidiary's stock following the spin-off (as was the case with Viacom, which canceled its Blockbuster spin-off for that reason).
  - On the other hand, note that the risk of cancellation is largely idiosyncratic and consequently should not be priced (because it can be diversified away by pursuing these investment strategies in a large number of carve-outs).

Short-sale constraints and the law of one price

Short-selling is an important, albeit imperfect arbitrage mechanism

Short-selling stock <u>is not</u> the mirror image of buying stocks, for institutional and legal reason

An investor who short-sells stocks sells stocks that he borrowed

At many exchanges, stocks cannot be shorted on downticks

Also, following IPO's (e.g., carve-outs), brokerage houses tend to agree not to lend stocks to short-sellers for a certain period of time (e.g., a couple of days)

Short sellers face a liquidity risk

Short seller might have to cover their shorts if they find themselves on the wrong side of the market (stop loss), for instance because they are <u>squeezed</u> by large shareholders

Also, the lender might decide to sell the stock, in which event the short-seller has to deliver.

The cost of shorting is reflected in the <u>interest rate rebate</u> short-sellers receive on the short-sale proceeds

The <u>lender</u> earns interest on the proceeds from the short-sale

The borrower (short-seller) receives part of this interest income as a rebate; the rebate functions as a price that equilibrates supply and demand in the securities lending market; the rebate might be negative.

The securities lending market (market for borrowing securities for short-selling) is fragmented with little transparency

- Parties interested in making a deal (borrowers, lenders) have to search for each other
- Stock investors that are most exuberant often have a very short average holding period, sometimes as short as a day, which makes it difficult to find shares to borrow.

The authors' Table 5 displays the short interest of the six carve-outs that the authors have identified as having a negative stub (at some point between the carve-out and the spin-off)

Note that the first column refers to the short interest of the parent

Table 5
Percent Short Interest

	First Month Parent	First Month Subsidiary	2nd Month Subsidiary	Peak Subsidiary
Creative/UBID	4.2	8.5	54.7	70.9
HNC/Retek	7.5	19.8	37.4	53.4
Daisytek/PFSWeb	1.6	17.7	48.6	63.7
Metamor/Xpedior	4.9	17.2	24.6	26.8
3Com/Palm	2.6	19.4	44.9	147.6
Methode/Stratos	1.5	31.8	50.3	114.7
Average	3.7	19.1	43.4	79.5
Difference from previous column		15.3	24.3	36.1
T-stat		4.4	4.5	2.3

Short interest calculated as percent of parent shares outstanding or subsidiary shares trading. The level of short interest comes from NASD, and is on or prior to the 15th calendar day of the month. The shares outstanding of the parent are from CRSP and the shares issued in the IPO are from company SEC filings. "First month" is the first observed short interest after the IPO, and "2nd month" is one month later. "Peak" is the highest level between the IPO date and the distribution date.

The table shows that …

… the subsidiary had considerably higher short interest than the parent

… the short-interest on the subsidiaries built up sluggishly

Remember the authors' Figure 3 of the—albeit not representative—time path of the stub value of Palm; the figure shows that the stub value was lowest at the IPO, then increased steadily (albeit not monotonically) over time before turning positive on the day of the IRS approval

At the peak short sales of Palms amounted to 147.6 percent of the floating shares; at that time, the stub value of Palm was already positive.

Note that sluggishness in the buildup offers only weak, indirect evidence for short-selling constraints.

To establish stronger (albeit also indirect) evidence for short-selling constraints the authors look for violation of the law of one price between shorts and synthetic shorts, the latter consisting of options and lending (or borrowing) at the risk-free rate

By definition, a synthetic short is a portfolio that—for all possible states of nature—offers the same final wealth as the genuine short

A synthetic short can be constructed by simultaneously …

… purchasing an at-the-money put and

… writing an at-the-money call and

… borrowing (at the risk-free rate) to square the balance.

The positions must be of the same time to expiration (maturity) as the genuine short.

The authors' Table 6 compares the prices of shorts and synthetic shorts for three different expiration dates (option expiration dates of May, August and November, 2000, respectively); the authors ignore the costs of shorting

Table 6
Palm options on 3/17/00

LIBOR	
three month	6.21
six month	6.41

Stock prices	
Palm	55.25
3Com	68

Options Prices

	--- Call ---		--- Put ---		Synthetic	Percent	Synthetic	Percent
	Bid	Ask	Bid	Ask	Short	Deviation	Long	Deviation
May 55	5.75	7.25	10.625	12.625	47.55	-14	51.05	-8
August 55	9.25	10.75	17.25	19.25	43.57	-21	47.07	-15
November 55	10	11.5	21.625	23.625	39.12	-29	42.62	-23

May options expire 5/20/00. August options expire 08/19/00, November options expire 11/18/00. A synthetic short position buys a put (at the ask price), sells a call (at the bid price), and borrows the present value of the strike price. A synthetic long position sells a put (at the bid price), buys a call (at the ask price), and lends the present value of the strike price. We discount May cash flows by three month LIBOR and August and November cash flows by six month LIBOR. Source of options price data: CBOE. Source of LIBOR: Datastream

The stock price of Palm that day runs at $55.25 (4 p.m. ET)

Selling short one share of Palm fetches $55.25, while doing it synthetically in the options market for the March-November holding period, for instance, fetches only $39.12

> The percentage deviation amounts to 29 percent, which means that the holding costs of short-selling would have to run at 29 percent or an annual rate of 147 percent to reconcile the price difference with the law of one price
>
> Given that estimates for the costs of shorting Palm during this time run at 30 percent, the authors conclude that there was a violation in the law of one price, which was caused by short-selling demand that the securities lending market could not meet.

The table also shows that buying a synthetic long (by simultaneously buying a call and writing a put, with the balance financed through borrowing at the risk-free rate) is cheaper than buying the actual stock!

The strongest evidence for a violation of the law of one price, however, is established by the fact that—as shown in the table—the price of (American-style) puts on Palm exceed the price of calls

For European-style options, the following put-call parity can be established (by devising synthetic calls and puts and equating their prices to the prices of genuine calls and puts):

$$\text{put}^{\text{European}} = \text{call}^{\text{European}} - \text{stock} + \frac{\text{strike price}}{1+r}$$

where r is the risk-free rate.

It can be shown that it is never optimal to exercise an American-style call on a non-dividend-paying stock (such as Palm)

Consequently, Palm calls are priced like European-style calls.

On the other hand, American-style puts on stocks always are worth more than corresponding European-style puts, because the right on early exercise is valuable

At any point in time, the following must hold:

$$\text{put}^{\text{European}} < \text{put}^{\text{American}} < \text{put}^{\text{European}} + r \cdot \frac{\text{strike price}}{1+r}$$

When inserting ($\text{put}^{\text{European}} > \text{put}^{\text{American}} - r \cdot \frac{\text{strike price}}{1+r}$) into the put-call parity for European options, we obtain:

$$\text{call} - \text{stock} + \frac{\text{strike price}}{1+r} > \text{put}^{\text{American}} - \cdot \frac{\text{strike price}}{1+r}$$

$$\Leftrightarrow \text{call} - \text{stock} + \text{strike price} > \text{put}^{\text{American}}$$

Thus the following inequality must hold for exchange-traded at-the-money options on Palm (which are American-style):

$$\text{call} > \text{put}$$

As shown in the authors' Table 6 (above), Palm puts were about twice as expensive as Palm calls, a clear violation of the law of one price.

Conclusion

- It takes many white swans to make us belief that all swans are white, but it takes only one black swan to destroy this belief
  - The fact that there have been instances of blatant violations of the law of one price in the stock and options markets casts doubt on rational asset pricing
  - Clearly, violations of the law of one price do not imply that markets are inefficient in the sense of money being on the table; also, there is the joint hypothesis problem
  - Yet, what is a concept of market efficiency worth that is consistent with such blatant mispricing?

[Blank page]

Growth-Optimal Investing

20. Stock Market Valuation

References:

Campbell, John Y., and Robert J. Shiller (2001) "Valuation Ratios and the Long-Run Stock Market Outlook: An Update," National Bureau of Economic Research, Working Paper 8221, http://nber.org/papers/w8221.

Shiller, Robert J. (1981) "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review* 71, 421-436.

The problem of stock market valuation

The problem of stock market valuation can be broken down into two sub-problems

Valuation of the stock market relative to <u>intrinsic value</u>

The intrinsic value of the stock market (as measured by an index, such as the S&P 500) is the present value of dividend payments plus share repurchases, adjusted for stock splits and the like

The intrinsic value is an <u>ex-ante concept</u> because it pertains to a yet unknown stream of payments.

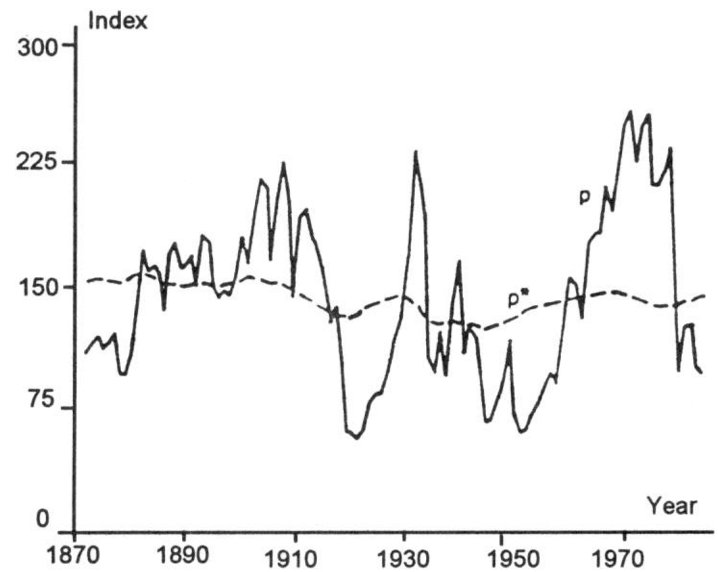
Valuation of the stock market relative to <u>historical valuations</u>

Clearly, if the stock market valuation is <u>unbiased</u>, fluctuations around the intrinsic value cancel out over time and the two sub-problems are identical.

Stock market valuation and intrinsic value

- In the chapter “Excessive Stock Market Volatility” we discussed a concept of calculating the intrinsic value of the stock market <u>ex post</u> from the realized dividend stream, devised by Shiller (1981)
  - As the discount rate, Shiller uses the real (i.e., inflation-adjusted) risk-free rate, which is an ex-post concept also in that it is calculated from realized (rather than expected) inflation rates
    - The figure below shows that the stock market, as represented by the S&P 500 Stock Price Index, exhibits pronounced swings around its intrinsic value

Figure 1



Note: Real Standard and Poor's Composite Stock Price Index (solid line p) and *ex post* rational price (dotted line p^*), 1871–1979, both detrended by dividing a long-run exponential growth factor. The variable p^* is the present value of actual subsequent real detrended dividends, subject to an assumption about the present value in 1979 of dividends thereafter.

Source: Shiller, Robert J. (1981) “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?” *American Economic Review* 71, 421-436.

Historical stock market valuation

Given that it is impossible to calculate the intrinsic value of the stock market directly, stock market valuation rests largely on historical comparison

The three most important concepts of stock market valuation are …

… P/E ratios (price-to-earnings ratio)

… the market-to-book ratio

… the “Fed model.”

P/E ratio (price-to-earnings ratio)

Published P/E ratios might …

… be four-quarter trailing P/E ratios

… might look back three quarters and use an estimate for the current quarter

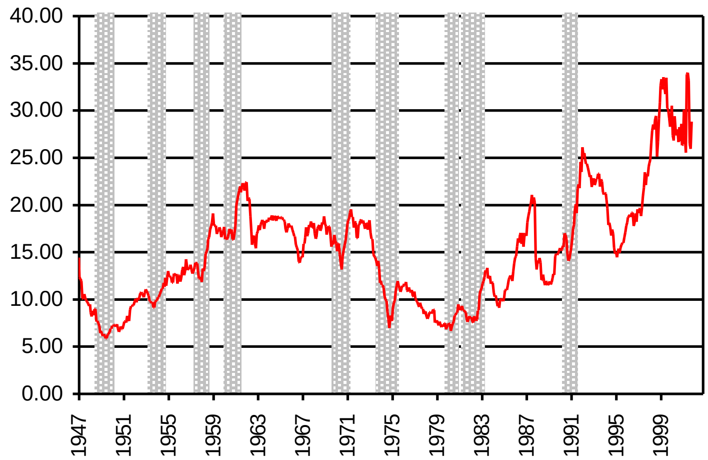
… be four-quarter forward-looking P/E ratios.

Also, there are alternative earnings concepts that are used for calculating P/E ratios

Operating earnings (after tax) is net income from on-going operations, which excludes non-recurring items, such as extraordinary items, earnings from discontinued operations, some special items, and gains or losses on sales of assets

Ordinary earnings (after tax) is net income from continuing operations calculated using generally accepted accounting principles (GAAP), which includes all current revenue and expenses.

Figure: S&P Composite 500, P/E Ratio, 4-Quarter Trailing Earnings



Note: Monthly data; first observation: January 1947; last observation: September 2001; median value: 14.88; shaded bars indicate recession periods. Source: Haver Analytics.

Note that the P/E ratio should not be constant

P/E ratios should (and indeed do) rise when earnings decline (e.g., when the economy goes into recession) and should decline when earnings improve

This is because temporary changes in earnings have little bearing on the value of an infinitely lived stock market.

Market-to-book ratio of equity

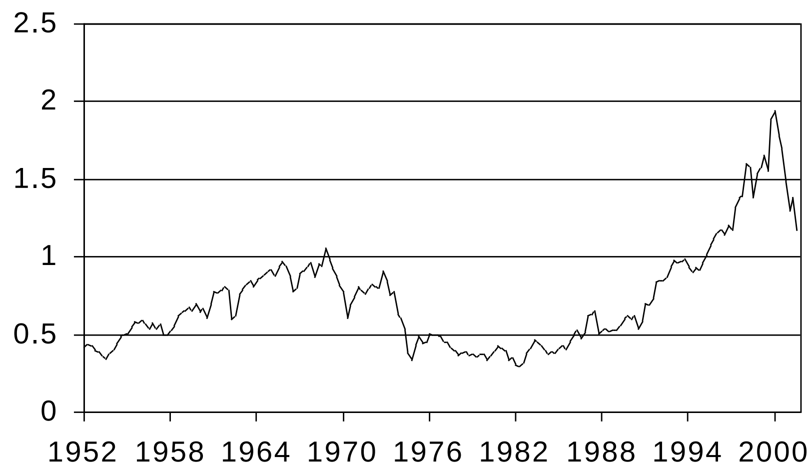
The market-to-book ratio of the corporate sector is typically measured as the ratio of market value of equities to financial net worth

A high market-to-book ratio of the corporate sector (relative to historical average) might indicate an increased importance of technical and organizational skills that are not reflected on the balance sheet

Alternatively, a comparatively high market-to-book ratio of the corporate sector might be viewed as a sign of excessive valuation.

The following figure shows the market-to-book ratio of the U.S. corporate sector

Nonfarm Nonfinancial Corporate Business: Market Value/Net Worth



Note: Quarterly observations (percent); first observation: 1952:1; last observation: 2001:3; source: Federal Reserve Board.

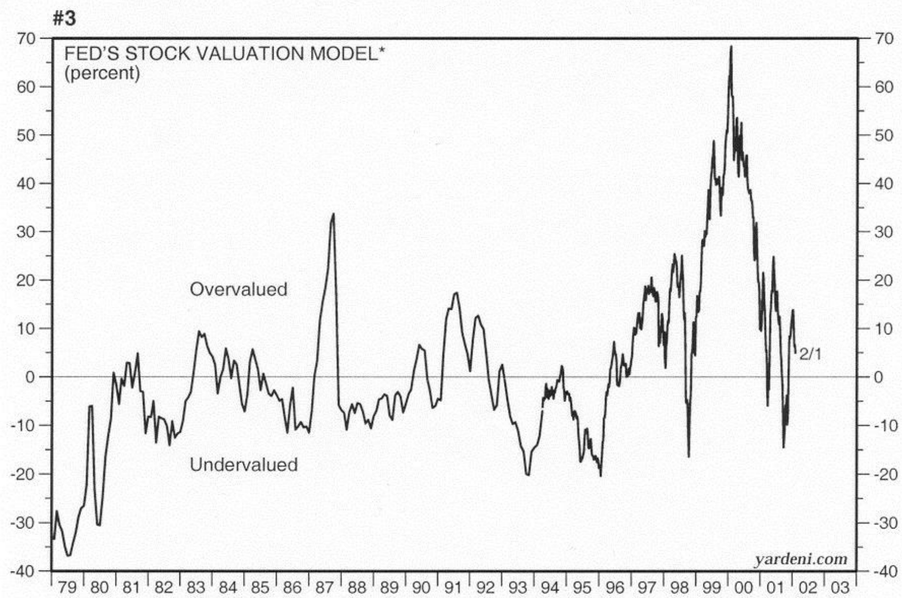
Related to the market-to-book ratio is Tobin's q

- Tobin's q is the ratio of the market value of the corporate sector (equity plus debt) to the replacement costs of assets
- There is an average and a marginal Tobin's q
- The corporate sector should expand if the added market value of new projects exceeds the costs of the newly employed assets, i.e., if marginal Tobin's q exceeds unity
- Typically, published data on Tobin's q typically refers to average (rather than marginal) Tobin's q
  - For traded companies, the market value of equity can easily be measured by the stock market capitalization
  - The market value of debt—save traded bonds—can be approximated by the book value of debt
  - The replacement costs of assets, on the other hand, are hard to measure.
    - Given that Tobin's q is hard to measure, researchers often use the market-to-book ratio of equity as a proxy.
- A high value of Tobin's q—as approximated by the market-to-book ratio in the above figure—might be indicative of a very favorable investment opportunity, or as a gross overvaluation of the U.S. corporate sector.

The Fed model rests on the relation between the earnings yield of the stock market (index) and the yield to maturity of government bonds

- The Fed model originates from remarks by Fed Chairman Alan Greenspan in the July 1997 Monetary Policy Report to the Congress (http://www.federalreserve.gov/boarddocs/hh/1997/july/FullReport.pdf)
  - One of the strongest proponents of the Fed model (in the business world) is Ed Yardeni, http://www.yardeni.com
  - The Fed model divides the 12-month forward earnings yield of the S&P 500 by the yield of the 10-year Treasury notes
    - Stocks are real assets and subject to fundamental risk (which means that the price at which they can be sold at retirement is uncertain)
    - Treasury notes, on the other hand, are nominal assets and are not subject to fundamental risk (i.e., their value at the date of maturity is known; there is basically no default risk)
    - Given the very different nature of the two assets, there is no solid theoretical foundation for the Fed model.
  - In spite of being used by Chairman Greenspan in a testimony to the Congress, the Fed model is by no means an official Federal Reserve stock market valuation model.

Figure: Stock market valuation using the Fed model

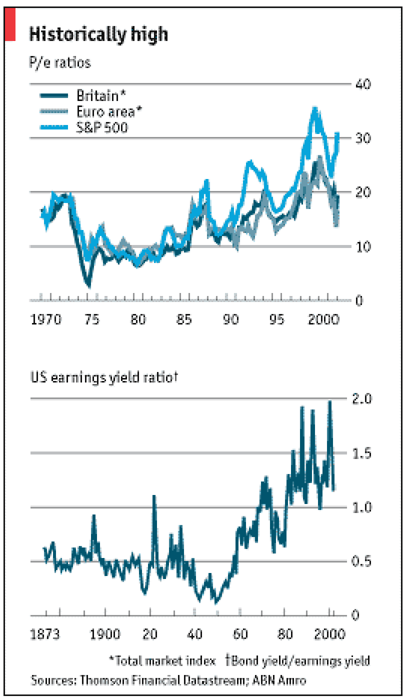


* Ratio of S&P 500 Index to its Fair-Value (52-week forward consensus expected S&P 500 operating earnings per share divided by the 10-year US Treasury bond yield) minus 100. Monthly through April 1994, weekly thereafter.
Source: Thomson Financial

Source: “Asset Valuation and Allocation Models,” February 4, 2002, p. 7, at http://www.yardeni.com/public/val.pdf, Valuation—US Stock Valuation Models—Current Week.

Conclusion

- Stock market valuation rests almost entirely upon historical averages
  - P/E and MTB ratios on one hand and the Fed model on the other hand, arrive at different conclusions about the current stock market valuation
    - According to the Fed model, the stock market (S&P 500) is currently (early February 2002) trading at around fair value
    - Based on P/E ratios, the U.S. stock market appears to be 100 percent overvalued
    - The conclusions offered by the Fed model can be reconciled with the conclusions offered by the P/E ratio approach if a longer time perspective is taken (see panel “US earnings yield ratio”)
      - In other words, both the Fed model and the P/E ratio approach indicate that the U.S. stock market is about 100 percent overvalued (as of November 2001)



Source: *The Economist*, November 22, 2001.

[Blank page]

Growth-Optimal Investing

21. Random Walk versus Mean Reversion

References:

Buffett, Warren (2001) “Warren Buffett on the Stock Market,” with Carol Loomis, *Fortune Investor’s Guide 2002*, December 10, 2001, http://www.fortune.com.

Campbell, John (2002) “Short-Sighted View of the Long Haul,” Financial Times, May 13, 2002, http://www.ft.com.

Campbell, John Y., and Robert J. Shiller (2001) “Valuation Ratios and the Long-Run Stock Market Outlook: An Update,” National Bureau of Economic Research, Working Paper 8221, http://nber.org/papers/w8221.

Knight, Frank H. (1921) “The Meaning of Risk and Uncertainty,” Part III, Chapter VII from *Risk, Uncertainty, and Profit*. Boston: Houghton and Mifflin. Online edition: http://www.econlib.org/library/Knight/knRUP.html.

Schmid, Frank A. (2002) “Stock Market Risk and Uncertainty,” *The Regional Economist*, July 2002, Federal Reserve Bank of St. Louis, http://www.stls.frb.org.

Schmid, Frank A. (2002) “Gambler’s Fallacy?” *Monetary Trends*, April 2002, Federal Reserve Bank of St. Louis, http://www.stls.frb.org.

Shiller, Robert J. (1981) “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?” *American Economic Review* 71, 421-436.

Shiller, Robert J. (2000) *Irrational Exuberance*, Princeton (NJ): Princeton University Press.

Siegel, Jeremy J. (1998) *Stocks for the Long Run: The Definitive Guide to Financial Market Returns and Long-Term Investment Strategies*, New York: McGraw-Hill; see also http://www.jeremysiegel.com.

Stock market investing and retirement

In the United States, employees have access to a variety of retirement savings vehicles, which give their owners the option of investing funds in the stock market

401(k) [or 403(b)], with a $15,000 maximum contribution for the year 2002

IRA or Roth IRA, each with a $3,000 maximum contribution for the year 2002

401(k) [403(b)] contributions are made bimonthly directly from the paycheck, while IRA and Roth IRA contributions are typically made once a year

The investment style of committing to a schedule of fixed contributions—as practiced in 401(k) [403(b)] plans—is called dollar-cost averaging

401(k) [403(b)] plans offer a variety of funds, among them typically a stock market (S&P 500) index fund.

Should the personal investor worry about stock market valuation and try to time the market?

Remember the noise trader from the chapter “Noise Trader Risk (Limits of Arbitrage I),” who exhibits the worst possible market timing

Also, remember the positive feedback trader from the chapter “Feedback Strategies (Bubbles),” who buys high and sells low

There is reason to believe that personal investors attempting to time the market wind up with a capital growth rate lower than a strict buy-and-hold stock market index portfolio.

Are <u>expected</u> (excess) stock market returns predictable?

- As discussed in the chapter "Predictability of Returns," there is reason to believe that there is a modicum of predictability of <u>expected</u> (excess) stock market returns
  - Note that predictability of expected (excess) returns does not imply predictability of actual returns
    - In particular, predictability of expected (excess) returns is no violation of (at least certain strands of) the efficient market paradigm because predictability of expected (excess) returns does not imply that there is money on the table
      - The day before Federal Reserve Chairman Alan Greenspan talked about "irrational exuberance" in the U.S. stock market, (http://www.federalreserve.gov/boarddocs/speeches/19961205.htm), the Wilshire 5000 Price Index—the broadest stock U.S. market index—stood at 7219.72 (December 4, 1996)
        - The index peaked at 14751.64 on March 24, 2000
        - An investor who would have shorted the market on Mr. Greenspan's advice most likely would have been broke by the time the stock market decline set in.
      - Remember that …
        - "Markets can remain irrational longer than you remain solvent" (John Maynard Keynes)
        - Chairman Greenspan might not have had the correct stock valuation model.

While aggressive investment strategies (e.g., short-selling) might lead to financial ruin (as demonstrated in the chapter "Judging Investment Strategies"), predictability of expected (excess) returns might allow investors to time the market on a multi-year basis

There are large, multi-year swings in investor sentiment, which manifest themselves in high P/E (price-to-earnings) ratios

Figure 1.2

Price-Earnings Ratio, 1881–2000

Price-earnings ratio, monthly, January 1881 to January 2000. Numerator: real (inflation-corrected) S&P Composite Stock Price Index, January. Denominator: moving average over preceding ten years of real S&P Composite earnings. Years of peaks are indicated.

Source: Shiller (2000, p. 8).

Note that Shiller's data are inflation-adjusted

Critics point out that comparing P/E ratios over long time horizons is futile because the sectoral composition of the S&P 500 changes over time—"as does the economy," one might counter this argument

High P/E ratios predict low future returns

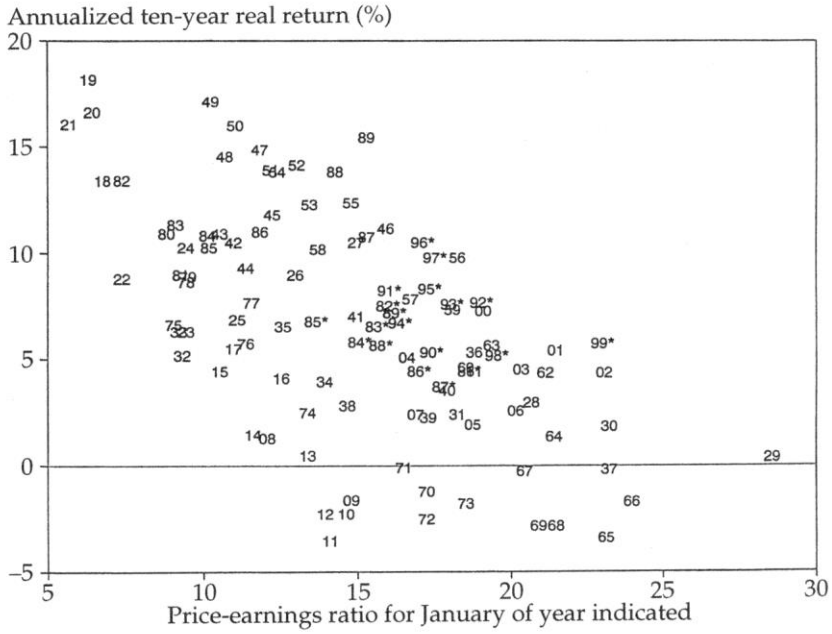


Figure 1.3

Price-Earnings Ratio as Predictor of Ten-Year Returns

Scatter diagram of annualized ten-year returns against price-earnings ratios. Horizontal axis shows the price-earnings ratio (as plotted in Figure 1.2) for January of the year indicated, dropping the 19 from twentieth-century years and dropping the 18 from nineteenth-century years and adding an asterisk (*). Vertical axis shows the geometric average real annual return per year on investing in the S&P Composite Index in January of the year shown, reinvesting dividends, and selling ten years later.

Source: Shiller (2000, p. 11).

Note that the return data plotted in the above figure are obtained from rolling ten-year windows and consequently are not independent observations.

Where is the stock market headed?

The stock market outlook depends on whether the log stock price index (e.g., the log of the S&P 500) follows a random walk, or whether there is mean reversion in the stock market

As mentioned in the chapter "Noise," stock prices (and stock market indexes, for that matter) are modeled as random walks

Remember that a variable following a random walk can wander off without bounds

The higher the frequency of the stock market data, the closer stock prices resemble a random walk.

For longer time intervals between data points (e.g., weeks, months, years), there is evidence for mean reversion in stock returns (see the chapters "Stock Market Overreaction") and stock market index returns (see the chapter "Excessive Volatility")

Remember that human behavior (e.g., performance, disposition) generally shows a great deal of mean reversion

Mean reversion in the stock market is introduced by investor sentiment (excessive optimism and excessive pessimism), and may take many years to run its course

Mean reversion is possible because arbitrageurs are not powerful enough to lean against the prevailing investor sentiment—due to the limits of arbitrage discussed in the chapters "Noise Trader Risk (Limits of Arbitrage I)" and "Professional Arbitrage (Limits of Arbitrage II)."

Judged by historical data, the stock market is the growth-optimal investment vehicle

The table below shows inflation-adjusted annualized growth rates of capital of a buy-and-hold portfolio (with dividends and capital gains reinvested) between the end of 1925 and the end of 2001 for the United States

Buy-and-hold Portfolio	Annualized Growth Rate 1926-2001 (Percent)
(Value-weighted index of) Large-company stocks	7.43
(Value-weighted index of) Small-company stocks	9.19
Long-term corporate bonds	2.63
Long-term government bonds	2.18
Intermediate-term government bonds	2.22
Treasury bills	0.73

Source: Ibbotson Associates, 2002, *Stocks, Bonds, Bills, and Inflation. 2002 Yearbook*. Chicago: Ibbotson Associates.

If the stock market follows a random walk, the growth-optimal investment strategy—based on historical data—is to be invested in the stock market at all times

This view was popularized by Jeremy Siegel in his 1994 book "Stocks for the Long Run" (2nd ed. 1998).

Remember from the chapter "Judging Investment Strategies" that growth-optimal investment is a survival-oriented concept, which takes due account of risk

On the other hand, if the stock market does not follow a random walk, being invested in the stock market at all times need not be growth-optimal.

There is also uncertainty, which is not captured by risk.

Stock market uncertainty

Uncertainty might manifest itself in changes of the data-generating process

When the data-generating process changes, historical data can no longer be relied on

Ironically, the data-generating process might change because we start incorporating historical data into our decision-making

The feedback of our own decision-making into the data-generating process—already discussed in Frank Knight (1921)—is known as the Lucas critique (see the chapter “Noise”)

Warren Buffet (2001) reports an interesting case of a change in the data-generating process in the stock market

Similar to Shiller (1981, “Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?” *American Economic Review* 71, 421-36), Buffet distinguishes between periods of (comparatively) high and (comparatively) low stock market valuation

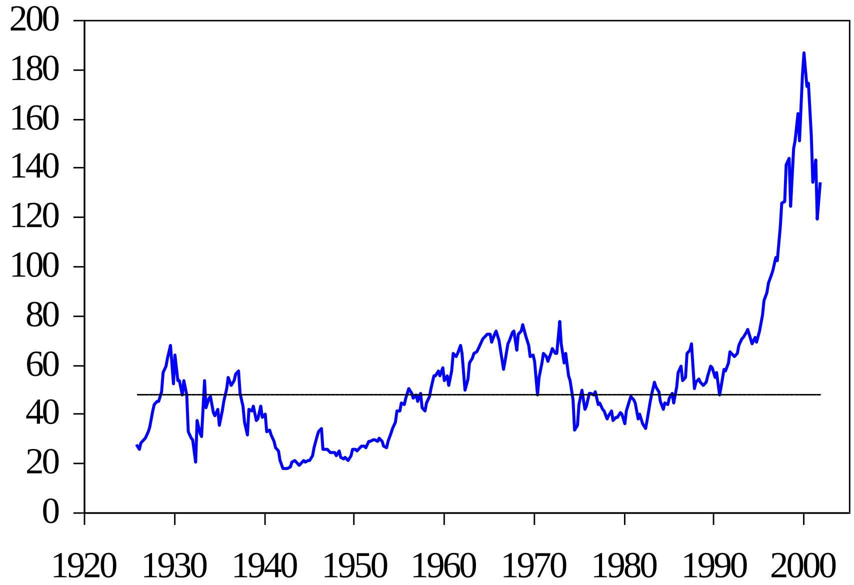
As shown by Shiller (1981), in the early 1920s stock market valuation was (comparatively) low, as measured by the inflation-adjusted discounted value of future dividends

The attractive valuation of stocks relative to bonds became a widely held belief after Edgar Lawrence Smith in 1924 published a book on stock market valuation, titled “Common Stocks as Long Term Investments”

- Smith argued that stocks do not only offer dividends, they also offer capital appreciation through retained earnings
  - The book, which received a complimentary review by John Maynard Keynes in 1925, gave cause to an unprecedented stock market appreciation
    - The inflation-adjusted average annual growth rate of a buy-and-hold investment in large-company stocks at the end of 1925 amounted to a staggering 32.13 percent at the end of 1928
      - On the other hand, over the next four years, this portfolio depreciated at an average annual rate of 22.93 percent, inflation-adjusted
      - Taken together, over the entire seven-year period the inflation-adjusted average annual growth rate of this portfolio came to a meager 1.11 percent.
    - Buy-and-hold portfolios in the allegedly unattractive long-term corporate and government bonds, on the other hand, grew at inflation-adjusted average annual rates of 10.18 and 9.83 percent, respectively!
    - We conclude from this episode that learning about the data-generating process of the stock market may feed back into the process
      - As investors’ behavior changes, the behavior of the stock market may change also.

Another history lesson in the making?

The following figure charts the stock market capitalization as a percentage of GNP



Note: Quarterly observations; first observation: 1925:4; last observation: 2001:4; median: 48.10; sources: Haver Analytics (GNP from 1948:1 to 2001:4), and Balke, Nathan S., and Robert J. Gordon (1986) "Appendix B: Historical Data," in: Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change*, Chicago: University of Chicago Press, 781-850 (GNP prior to 1984:1).

The figure above shows a sharp appreciation of the U.S. stock market relative to GNP, starting in 1994

Remember that in 1994, the first edition of Jeremy Siegel's "Stocks for the Long Run" was published

Did Jeremy Siegel's 1994 book have an effect on the American public similar to Lawrence Smith's 1924 book?

Mean reversion or dilution?

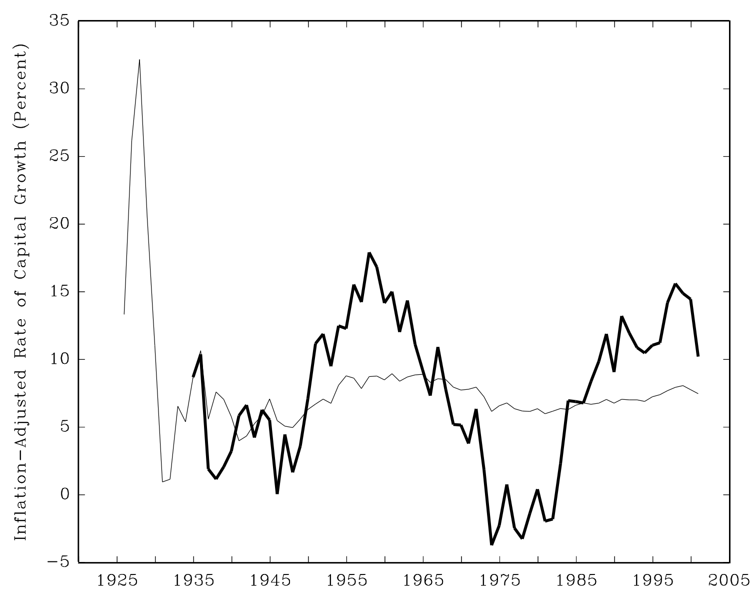
If the stock market mean-reverts, above-average growth rates of capital of buy-and-hold portfolios are corrected down the road

If there is mean reversion, past above-average rates of capital growth predict low future rates of capital growth, on average.

On the other hand, if the stock market follows a random walk, past above-average capital growth rates are simply diluted as the random process unfolds

If the market follows a random walk, past capital growth rates contain no information on future capital growth rates.

The following figure charts average rates of growth of a buy-and-hold portfolio in large-company stocks, inflation-adjusted



Note: Annual observations; first observation: 1926 (growth rate since 1925; thin line), 1935 (growth rate over last 10 year, thick line); last observation: 2001; source: Ibbotson Associates, 2002, *Stocks, Bonds, Bills, and Inflation. 2002 Yearbook*. Chicago: Ibbotson Associates.

Take your pick between dilution (random walk) and mean reversion!

[Blank page]