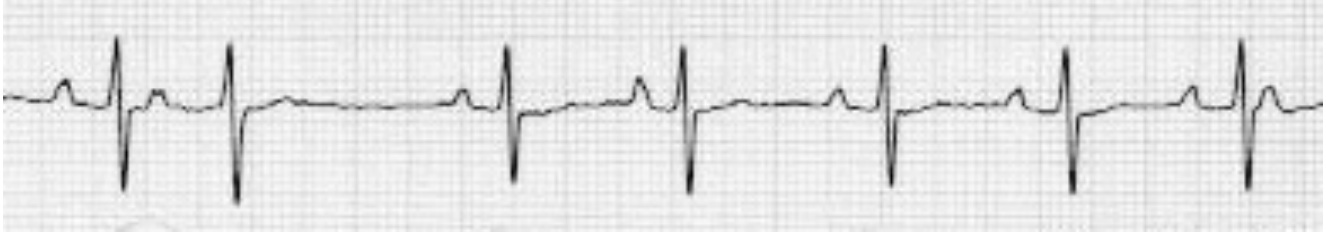
Information Analysis



"Whoa—way too much information!"

Introduction

- How can we judge the information content of an idea?
- More concretely:
  - Is this a signal we can profitably deploy on paper?
  - Is this a signal we can profitably deploy in practice?

Outline

- Basic information analysis
- The information horizon
- Backtesting
- Challenges

Basic Information Analysis

- Can we profitably deploy this signal on paper?
- Two-step analysis:
  - Turn information (signal) into portfolios
  - Analyze portfolio performance
- Apply the scientific method
  - Hypothesis testing
  - Controls
  - Statistics

Basics

- Minimum Requirements
  - A process
- Assumption
  - Historical analysis will provide some useful insight into future performance.
- Observation
  - Running information analysis is easy.
  - Finding valuable information is hard.

Step 1: Information into Portfolios

- Keep in mind an example as we go through this.
  - b/p ratios
- We have a historical record of the signal.
  - Monthly b/p ratios back 10 years
  - What universe of stocks?
- How do we turn signals into historical portfolios?

Signals into Portfolios

- Choices here limited only by creativity.
- *N*-tile analysis: (very popular with academics)
  - Rank stocks by signal.
  - Place first (1/*N*) in *N*-tile 1, etc.
  - Even this simple approach offers many choices.
- Factor portfolios
  - Controlled experiments

N-tile Analysis Choices

Quintiles or Deciles

- What is *N*?
- What universe of stocks?
- Do we divide stock universe
  - by number? — 100 per quintile
  - by cap?
- How do we weight stocks in each *N*-tile portfolio?
  - Equal-weight
  - Cap-weight
  - Something else?

Example:
S&P 500
Quintiles
(5 portfolios)

Let's say Mkt Cap of S&P 500 is 20 Trillion
4 Trillion into each quintile

Example : S&P 500
B/P ratios
Quintiles

Stock	B/P ratio	Mkt Cap
Stock 1	highest	
	⋮	
Stock 500	lowest	

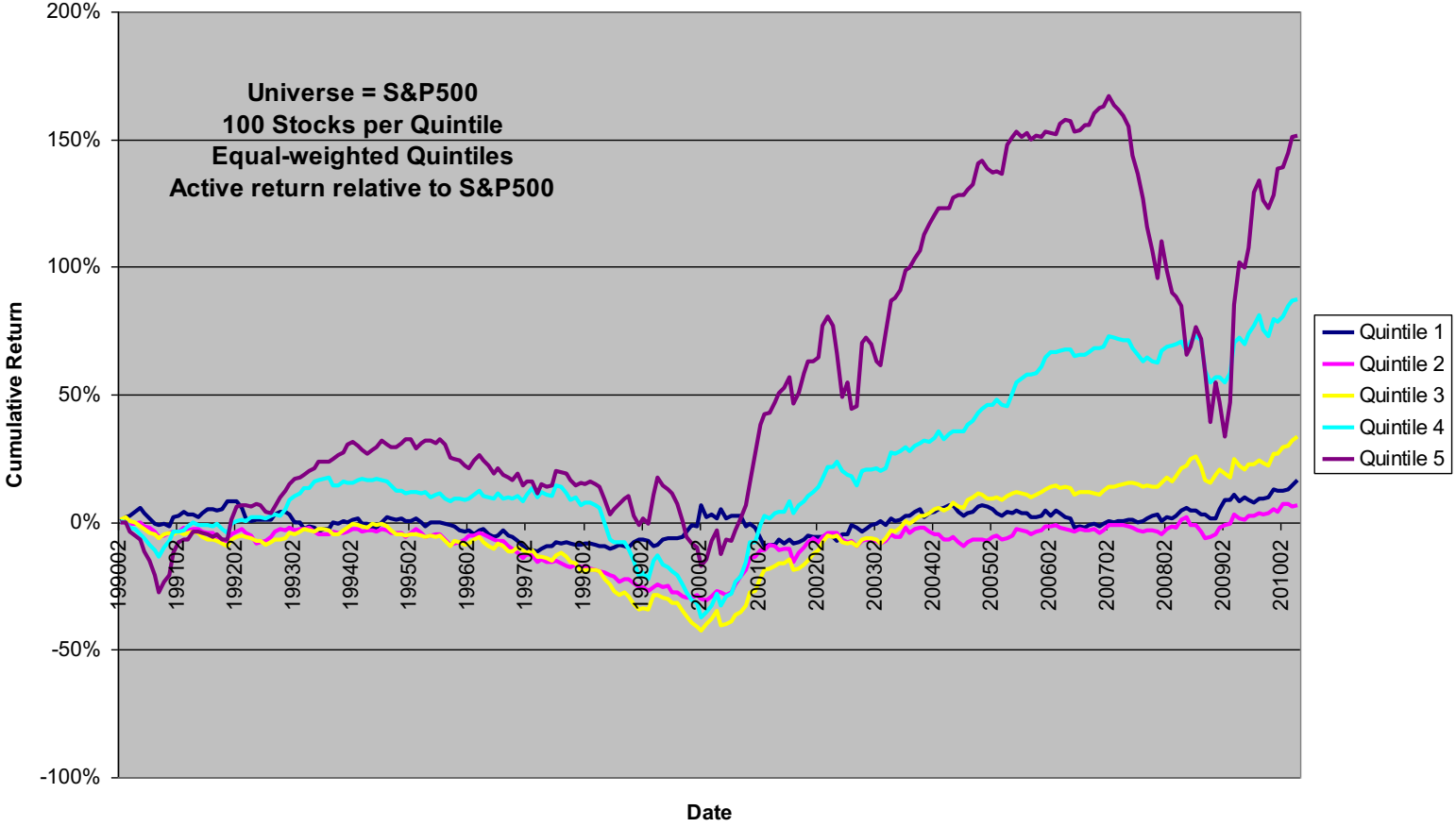
Approach 1:

100 Stocks in each quintile

Approach 2

$4 Trillion in Mkt Cap in each quintile

Book-to-Price Quintile Analysis



What are we looking for?

(1) Are the returns roughly ordered?
Quintile 5 > Quintile 4 > Quintile 3 > ...

(2) Academic Paper Standard
Decile 10 − Decile 1

Factor Portfolios

- Optimally construct portfolios
- Signal exposure one standard deviation above universe mean.
- Zero exposure to control variables
- Minimum risk.

Cross sectional
Portfolio Score = 1

Factor Portfolios

- Convert signal to cross-sectional score

$$z_n = \frac{g_n - Mean_{CS}\{g\}}{Std_{CS}\{g\}}$$

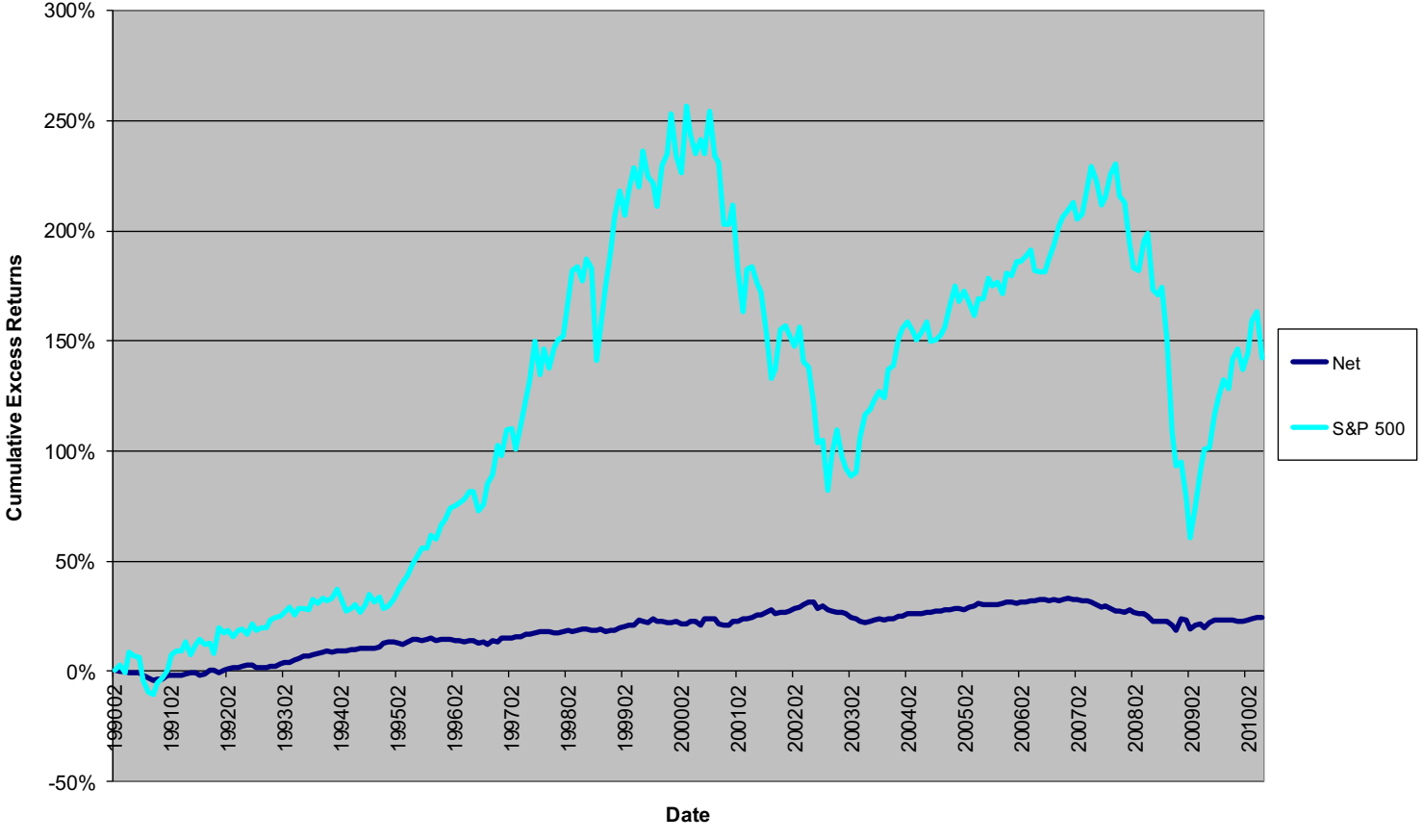
- Build optimal portfolio

$$Min\left\{\mathbf{h}^T \cdot \mathbf{V} \cdot \mathbf{h}\right\}$$

$$\text{Subject to } \mathbf{h}^T \cdot \mathbf{z} = 1$$

$$\mathbf{h}^T \cdot \mathbf{A} = 0$$

B/P Factor Portfolios



Control Variables

- Other factors potentially related to returns:
  - The market (beta)
  - Industries
  - Investment themes (e.g. size, value, momentum…)

Alternative Approach

Factor Model
$r = X \cdot b + u$

- Regression-based

$$\mathbf{r} = f_z \cdot \mathbf{z} + \mathbf{X} \cdot \mathbf{b} + \mathbf{u}$$

- The coefficient f_z is the return to a factor portfolio.
- Under certain circumstances these two approaches given identical results. More typically, they are reasonably close.

Problem with Factor Portfolios

- They are non-investible.
  - They take positions in every stock.
  - They have long and short positions.
  - We build them without regard for transactions costs.
    - So they are high turnover.
- They require a covariance matrix. (This is often a challenge for academics testing ideas. It isn't a problem for investment managers.
- Information analysis is the beginning of our analysis, not the last step.

Step 2: Performance Analysis

- We have now built historical portfolios for every period.
  - e.g. monthly b/p-based portfolios.
- We can now look at the performance of those portfolios over time.
  - Cumulative return plots ← I showed this to you in earlier slides
  - Alpha and beta of portfolios
  - t-statistic of the alpha
  - Information Ratio
  - Number of up versus down months
  - Performance in up and down markets
  - Turnover

Performance Analysis

- Key assumption:
  - Our signal is generating the performance.
  - We have controlled for all other relevant factors.
- We can examine such questions in detail

Example: Academic Analysis

① Signal ⟹ Decile Portfolios

② Look at returns: Decile 10 − Decile 1

$r_p(t)$ = Decile 10 − Decile 1

③ Regress:

$$r_p(t) = \alpha_p + \beta_1 \cdot r_{MKT}(t) + \beta_2 \cdot SMB + \beta_3 \cdot HML + \beta_4 \cdot MOMENTUM + \varepsilon$$

SMB – Size Portfolio

HML – Value Portfolio

MOM – Momentum Portfolio

Ken French website

④ Is α statistically significant

The Information Horizon

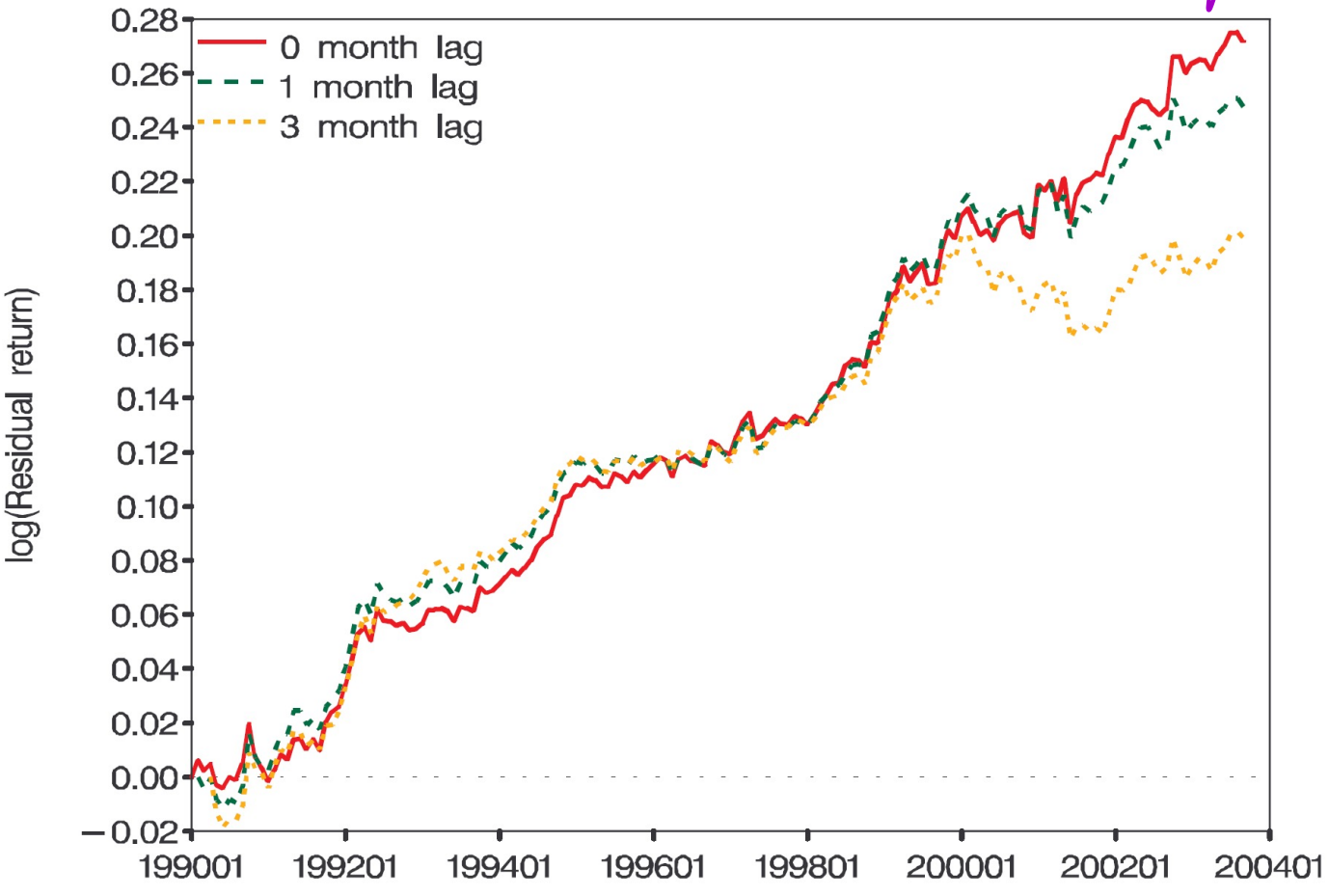
- The value of our information decays over time.
  - Information can grow stale.
- The Information Ratio will decay as information ages.
- The half-life of the information can quantify that effect.
- Longer horizons are typically desirable.

Measuring the Information Horizon

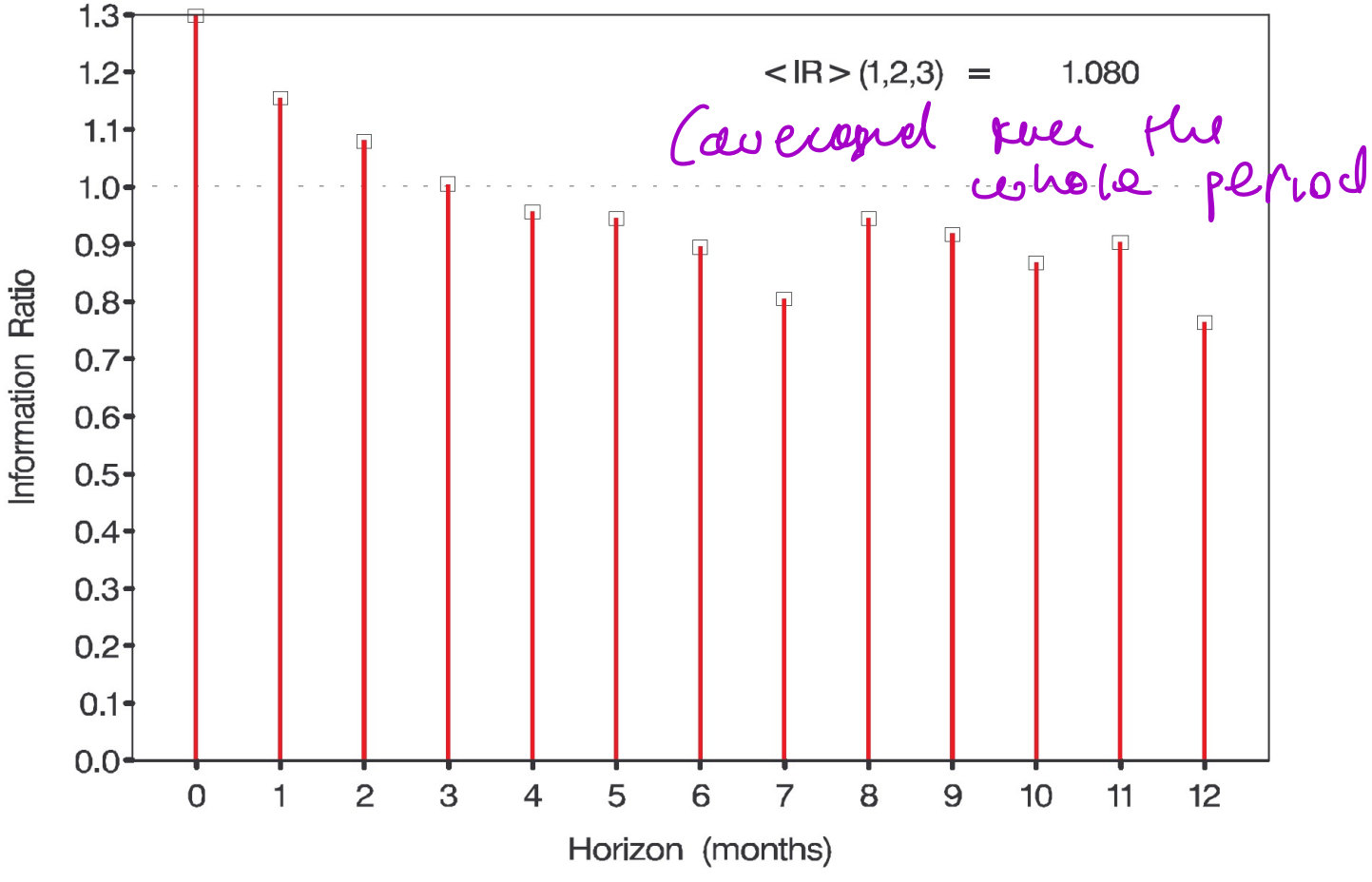
- Use time t information to build portfolios.
- Invest at time $t+s$.
- Monitor cumulative performance as a function of s.
- Monitor Information Ratios as a function of s.
- See examples

Cumulative Ret vs. Horizon (I)

Richard Sloan Accruals Signal



Horizon *IR* (I)

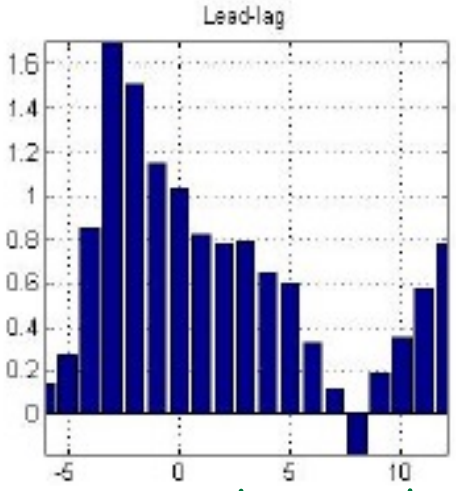


Averaged over the whole period

Signal II: Trending

Signal = $\sum_{t-1}^{t-12}$ monthly return

↳ e.g. momentum



t=0 Calculate past 12 month return.

Months Lead or Lag

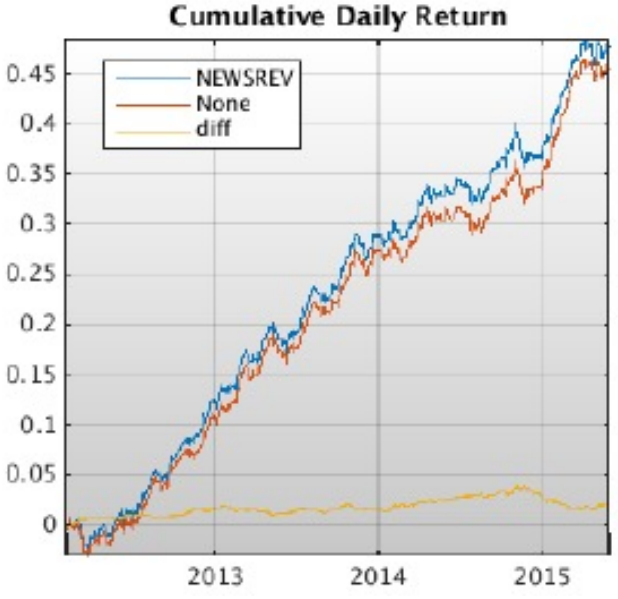
t=5 month lag

Monthly Backtests

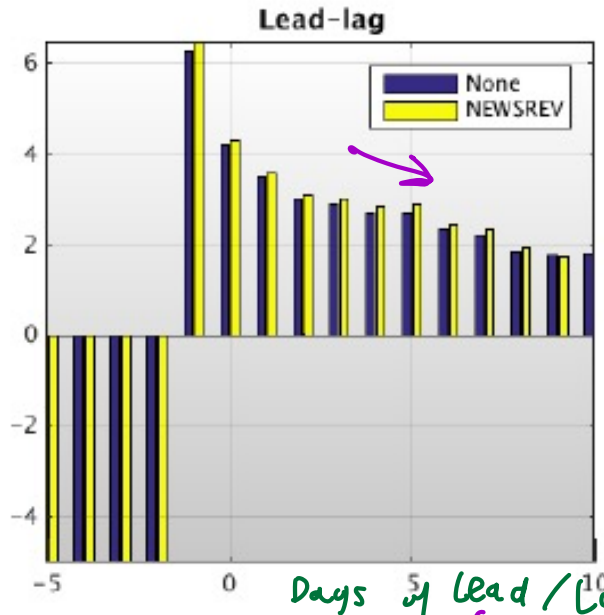
Build Factor Portfolio

Observe return

Signal III: Mean-Reverting



Daily Backtests



Signature of mean reversion leads are negative

Another Approach to Information Analysis: Event Studies

- Particular form of information analysis.
- So far we have considered cross-sectional approaches.
- But what if we wish to study events which occur at different times for different assets?
  - Earnings announcements.
  - Change in CEO
  - Etc.

Event Studies

- The event occurs at t=0, no matter what the calendar date.
- We need to carefully control how we aggregate post-event returns, given that they involve different stocks at different times.
- Solution: we control for risk factors and risk levels.

Event Study Returns

- Look at standardized outcomes:

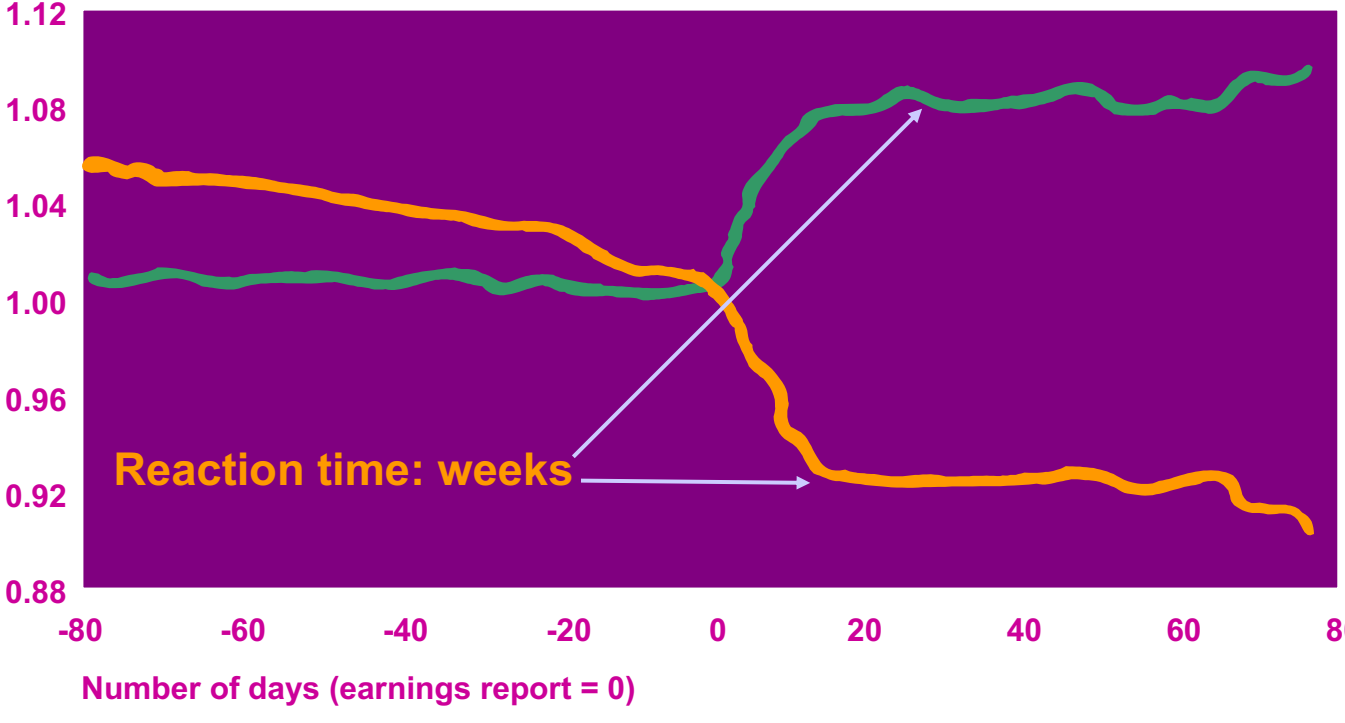
$$x_n(t) = \frac{\theta_n(t)}{\omega_n(t)}$$

- We can combine these. We can regress against other controls:

$$x_n(t) = a + \sum_{j=1}^{J} Y_j \cdot b_j + \varepsilon_n(t)$$

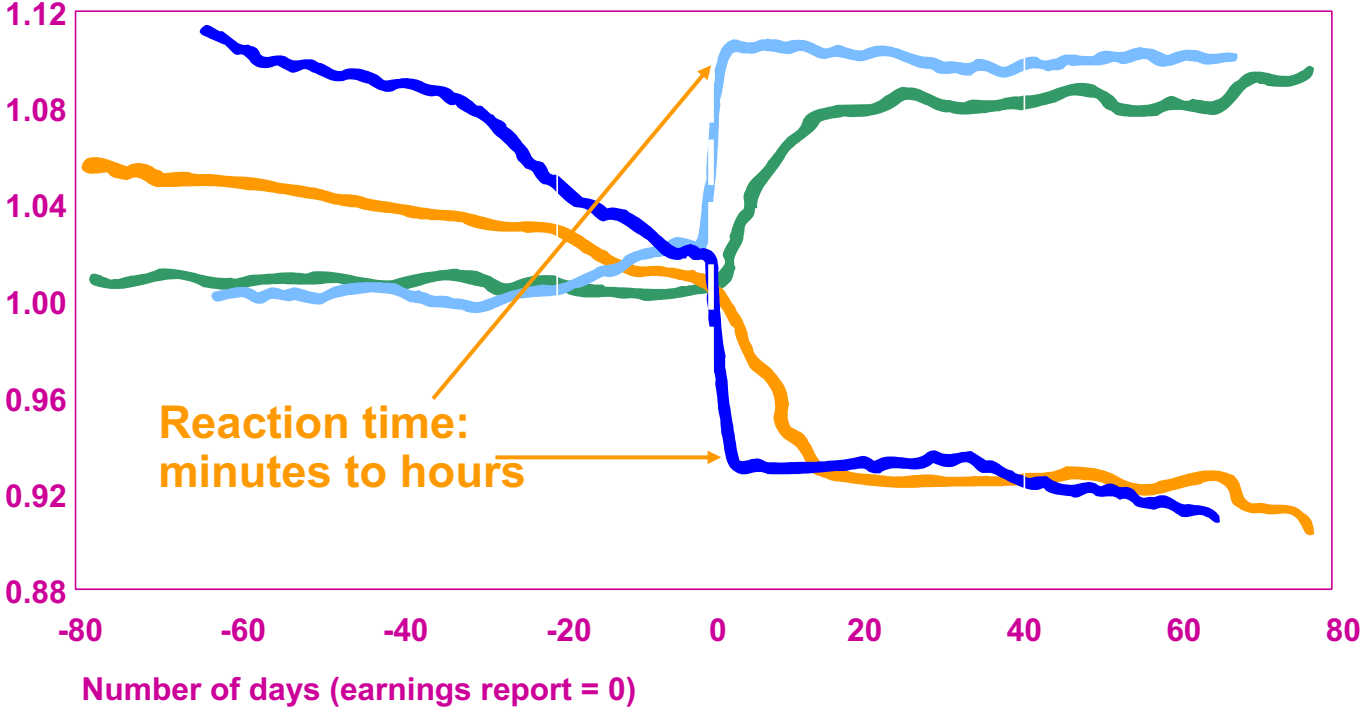
Example: Earnings Surprise

Earnings surprise effects then (1983-89)...



Evolution in markets

…and earnings surprise effects now (1995-98)



Backtesting

- More realistic version of information analysis.
- Investible portfolios.
- Account for constraints (like long-only) and transactions costs. These tend to lower performance relative to that observed in information analysis.

Backtesting, cont.

- Often backtests consider not only just lone signal performance, but the improvement in performance from adding new idea to existing alpha forecasts.
- Ultimately we will judge a signal based on whether it is:
  - Sensible
  - Predictive
  - Consistent
  - Additive

SPCA

- Ideally we will also run *ancillary* tests.
  - Other (non-return) implications of the hypothesis.
  - Remember Sloan's work on accruals.

What are the Challenges?

- Observation: In-sample results with 95% confidence disappoint far more than 5% of the time.
- Cautionary Tale
  - Norman Bloom: World's greatest dataminer.
  - Multiple lottery winners

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2×2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of "true relationships" to "no relationships" among those tested in the field. R is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2×2 table, one gets $\text{PPV} = (1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

Abbreviation: PPV, positive predictive value

John P. A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124

HIER

Harvard Institute of Economic Research

Discussion Paper Number 2122

Researcher Incentives and Empirical Methods

by

Edward L. Glaeser

October 2006

HARVARD UNIVERSITY
Cambridge, Massachusetts

Analyzing the Challenge

- Degree of difficulty:
  - Ex-ante fraction of ideas we investigate that are actually true.
  - Needles in haystacks vs fish in barrels.
- Probability of Type I error (false positive).
- Probability of Type II error (false negative).
- Bias (probability of false positive due to choice of data, analysis, and presentation)
- Number of tests.

Degree of Difficulty example

- Your doctor tests you for a rare (1/1000) disease. The test is 99% accurate (i.e. only 1% false positives).
- The test comes back positive. How likely is it that you have the disease?
- Your odds are about 1/11. ~ 9%
- Analysis: Out of 1,000 people, we expect one true positive. If we apply our 99% accurate test on the other 999 people, we expect 10 false positives. So testing 1,000 people will lead to 11 positives, only 1 true positive.

Degree of Difficulty

- Ex-ante ratio of positive to negative results (P/N):

$$R_{pn} \equiv \frac{pos}{neg}$$

- Hence:

$$\frac{pos}{pos+neg} = \frac{R_{pn}}{R_{pn}+1} \qquad \frac{neg}{pos+neg} = \frac{1}{R_{pn}+1}$$

Simple Analysis

- Assume we test *c* signals.
- We expect $\frac{c \cdot R_{pn}}{R_{pn}+1}$ are positive and $\frac{c}{R_{pn}+1}$ are negative.
- What fraction of those positives test positive?
- What fraction of those negatives test positive?
- What fraction of the positive test results are actually positive?

Type I and Type II Errors

- Fraction of false positives = f_{fp}
- Fraction of false negatives = f_{fn}
- Out of $\frac{c \cdot R_{pn}}{R_{pn}+1}$ actual positives, $c \cdot \left(\frac{R_{pn}}{R_{pn}+1} \right) \cdot \left(1 - f_{fn}\right)$ will test positive.
- Out of $\frac{c}{R_{pn}+1}$ actual negatives, $c \cdot \left(\frac{1}{R_{pn}+1} \right) \cdot f_{fp}$ will test positive.

Basic Analysis Set-Up

Out-of-Sample Results

In-Sample Results

	Value-adding	Valueless
Positive	$c \cdot \left(\frac{R_{pn}}{R_{pn}+1} \right) \cdot (1 - f_{fn})$	$c \cdot \left(\frac{1}{R_{pn}+1} \right) \cdot f_{fp}$
Negative	$c \cdot \left(\frac{R_{pn}}{R_{pn}+1} \right) \cdot f_{fn}$	$c \cdot \left(\frac{1}{R_{pn}+1} \right) \cdot (1 - f_{fp})$

$$Total\ True = c \cdot \left(\frac{R_{pn}}{R_{pn}+1} \right)$$

$$Total\ False = c \cdot \left(\frac{1}{R_{pn}+1} \right)$$

All test positive. What fraction of these are value-adding

= Positive Predictive Value

What fraction of positive test results add value out-of-sample?

This is the positive predictive value:

$$PPV = \frac{R_{pn} \cdot \left(1 - f_{fn}\right)}{R_{pn} \cdot \left(1 - f_{fn}\right) + f_{fp}}$$

You can see that as the ex-ante probability increases, and types I and II errors decrease, this fraction can approach 1. However it can also fall far below 1.

Embellishments

- Bias
  - Some fraction of analyses of negative results presented as positive due to bias. So statistical error, bias, or both will lead to negative results reported as positive.
- Multiple Tests
  - What if we test multiple variants of the signal until we find one variant that works?
  - False positives will increase because we will report the relationship as positive if even only 1 out of *N* tests shows up positive.

Application: Medical Research

Table 4. PPV of Research Findings for Various Combinations of Power (1 – β), Ratio of True to Not-True Relationships *(R)*, and Bias *(u)*

1 – β	*R*	*u*	Practical Example	PPV
0.80	1:1	0.10	Adequately powered RCT with little bias and 1:1 pre-study odds	0.85
0.95	2:1	0.30	Confirmatory meta-analysis of good-quality RCTs	0.85
0.80	1:3	0.40	Meta-analysis of small inconclusive studies	0.41
0.20	1:5	0.20	Underpowered, but well-performed phase I/II RCT	0.23
0.20	1:5	0.80	Underpowered, poorly performed phase I/II RCT	0.17
0.80	1:10	0.30	Adequately powered exploratory epidemiological study	0.20
0.20	1:10	0.30	Underpowered exploratory epidemiological study	0.12
0.20	1:1,000	0.80	Discovery-oriented exploratory research with massive testing	0.0010
0.20	1:1,000	0.20	As in previous example, but with more limited bias (more standardized)	0.0015

The estimated PPVs (positive predictive values) are derived assuming $\alpha = 0.05$ for a single study.
RCT, randomized controlled trial.
DOI: 10.1371/journal.pmed.0020124.t004

(Example of discovery-oriented exploratory research: testing 30,000 genes when perhaps 30 are the true culprits.)

Benedict Carey, NY Times, 8/27/15: More than half of psychological papers are not reproducible. Initiative to replicate findings of 100 prominent studies ("The Reproducibility Project") raises further questions about the health of the discipline.

Ioannidis Observations on Medical Research

- Research findings are less likely to be true:
  - The smaller the study.
  - The smaller the effect size.
  - The greater the flexibility in designs, definitions, analyses.
  - The greater the financial interests.
  - The hotter the field of study
- "Finally, …before running an experiment, investigators should consider what they believe the chances are that they are testing a true relationship."

Application: Financial Research

- Not medical research. We are not looking for immutable laws of nature, but rather relationships we hope will work for the next year or so. We fully expect most alpha signals (in particular alpha factors not related to risk premia) to stop working as the market discovers them.

Observations from Glaeser

- Researcher initiative is the norm, not the exception
  - Researchers will optimize to try to find high correlations
- The optimal amount of data-mining isn't zero.
  - This activity produces knowledge, though we must adjust our statistical techniques.
- Research occurs in a market where competition and replication greatly matter.
- Changes in technology generally decrease the costs of running tests, and increase the availability of potential explanatory variables.
- Methodological complexity offers researchers more degrees of freedom and increases the cost to competitors reproducing results. Skepticism applied to new more complex methods may be appropriate.
- Data collection and cleaning offer easy opportunities for improving statistical significance. Be more skeptical of analysts who produce and clean their own data.

Application: Signal Research

- Degree of difficulty:
  - R_{pn}<~0.1 for scattershot datamining.
  - R_{pn}~0.5 once we require sensibility.

Application: Signal Research

- Type I Errors: 5% typically.

- The number of false positives is higher than that if we are testing for success over the next 12 months.

What Fraction of Signal Research Results add value out-of-sample?

Research Environment	f_{fp}	f_{fn}	**Bias**	N	R_{pn}	**PPV**
Scattershot datamining	0.05	0.01	0.1	20	0.1	**9.5%**
No SPCA* process	0.05	0.05	0.2	10	0.15	**13.8%**
Sensibility	0.05	0.05	0.2	3	0.5	**47.1%**
SPCA, Ancillary Testing	0.01	0.05	0.05	3	0.5	**74.8%**

*SPCA stands for “Sensible, Predictive, Consistent, and Additive”