# And Still More Factors and More Factors

Matthew Rothman

MIT Sloan School of Management

**Spring 2023**

# Earnings Management

# Earnings Management

Hypothesis: Managers will ”manipulate” their earnings numbers to make the performance of the company look better than it really is

- Why might this do this?
  - Career Concerns: (i) CEO/CFO don’t want to be fired; (ii) CEO/CFO want to be hired at an even bigger company.
  - CEO/CFO want a raise
  - They want the stock price to go up as their deferred compensation is in stock
  - Customers don’t want to buy products of companies that might not exist
  - Debt covenants might have performance metrics
  - Employees might leave if they think the firm is in trouble
  - Need to raise new capital in form of debt or equity and want continued access to trade credit
  - They want to juice the price of the stock before M&A
  - Ego
  - Rat Race Equilibrium

# Earnings Management

How might they do it?

1. Big Bath

   a. You take one-time charges against income, reducing asset values, resulting in lower income this year but leading to larger income in the future years

      Example: Inventory on the books is written down this year from $100 to $50. This results is a net loss of $50 this year. But when the item is sold in the future for $75, it takes what would have been a $25 loss and turns it into $50 profit.

   b. Can be in large write-offs, asset impairment charges, restructuring charges, and litigation losses. The are generally reflected in special items in the financial statements

   c. Why would they do this?

      – New Management has come in and wants to show they are ”cleaning house” and “goosing profits” to look in future years.

      – They will have big losses already this year, so just “take the hit now”.

# Earnings Management

How might they do it?

2. Smoothing Earnings

   a. There are two broad types of earnings: cash flows and accruals. Accruals are the non-cash component of earnings that help align expenses and revenues across time.

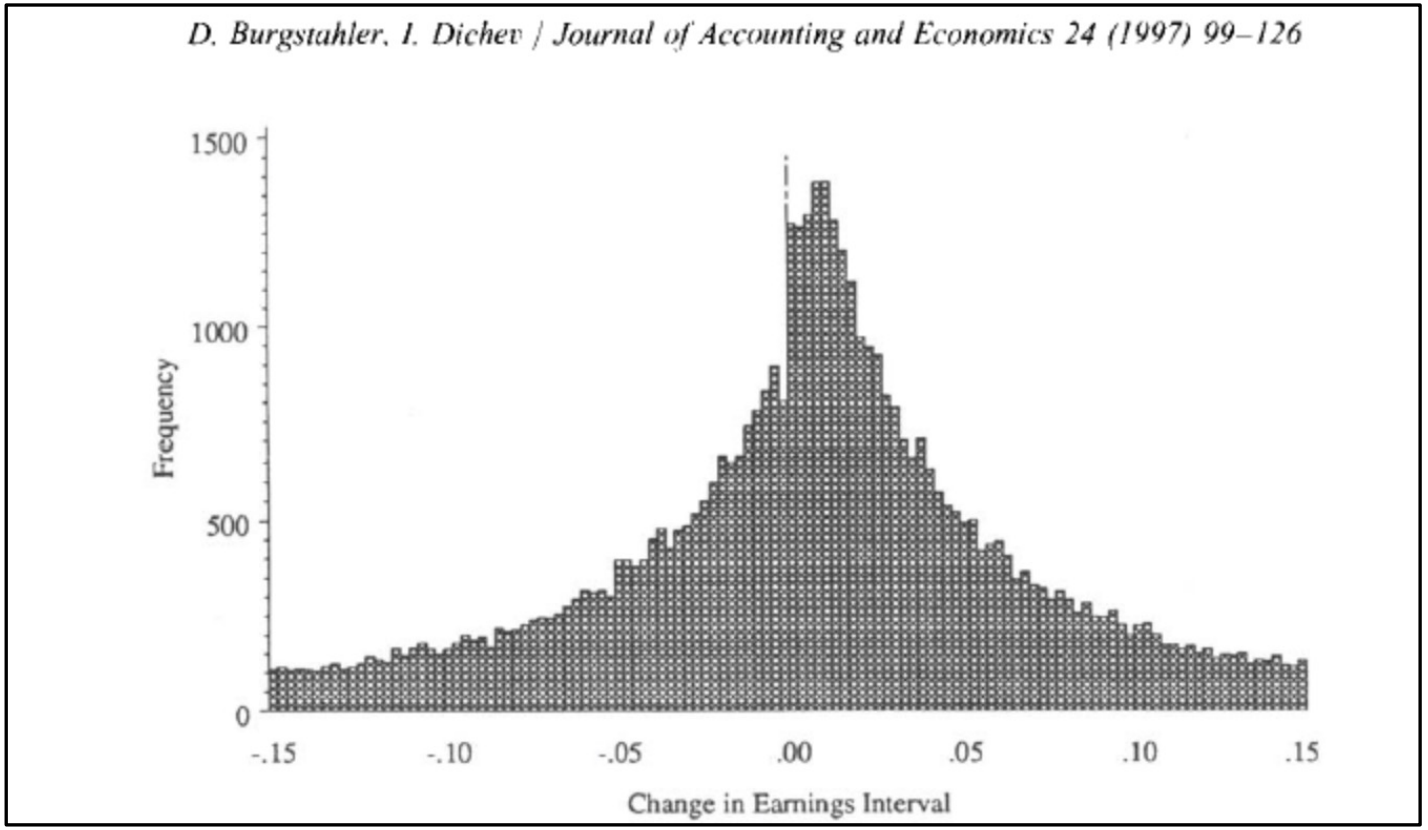
   In order to meet or beat earnings expectations and/or to provide the appearance of consistent growth, companies may manage there reported earnings number through the deliberate use of increasing the accrual component of earnings.

   They would do this for the reasons previously discussed.

   **Question: how might one detect either of these two forms of earnings management?**

# Earnings Management: Evidence

Earnings have a notable discontinuity around changes from prior year's earnings

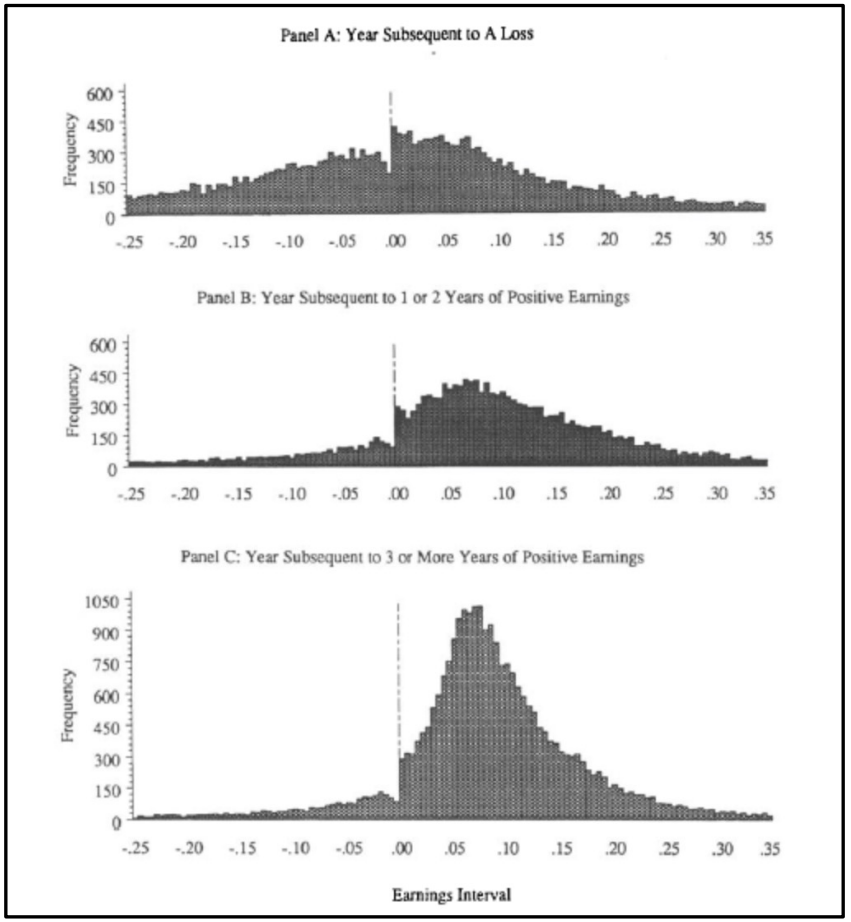



Evidence from 1976-1994, for over 77k firms finds a high significant deviation around "zero" measured as the change from prior year's earnings

Source: Burgstahler and Dichev "Earnings Management to Avoid Earnings Decreases and Losses" (1997), *Journal of Accounting and Economics*

# Earnings Management: Evidence

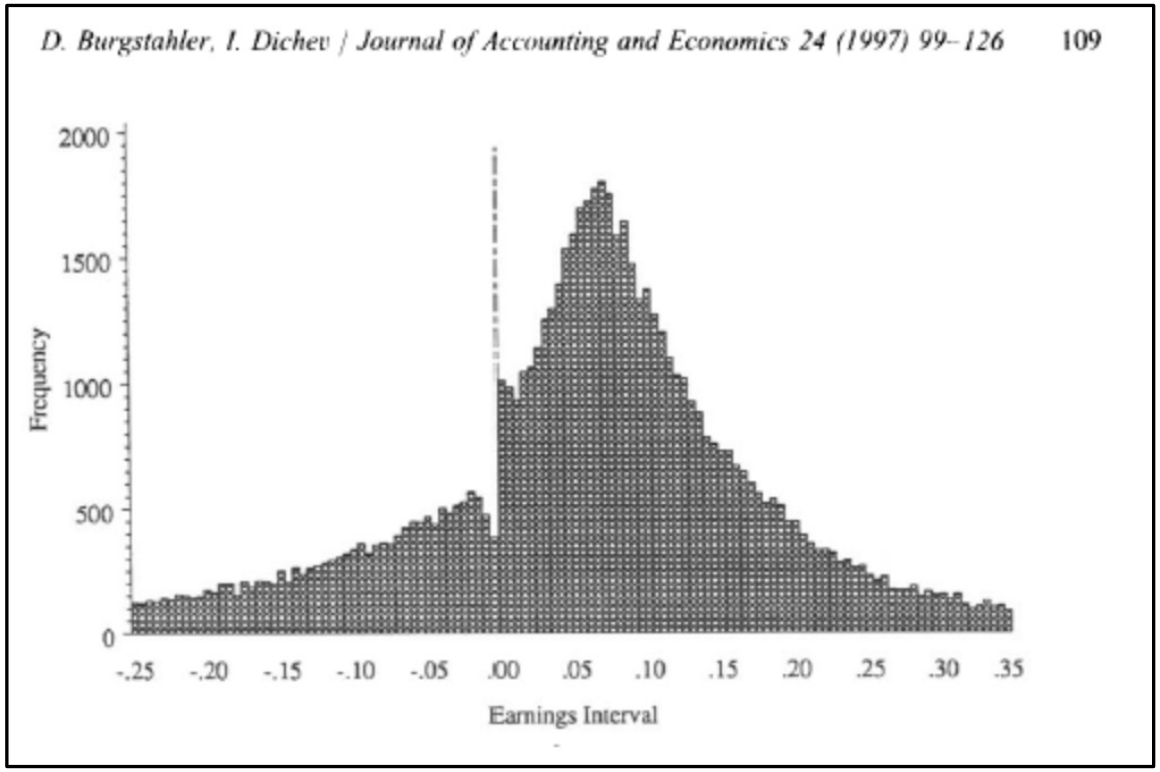
Earnings have a notable discontinuity around changes from prior year's earnings




The longer and stronger the trend, particularly positive, the more management doesn't want to show a break …. Or the more they have managed in the past, the more likely they are to be managing today …. Habitual Earnings Managers!

# Earnings Management: Evidence

Earnings have a notable discontinuity around zero – that is, firms want to avoid showing an earnings loss

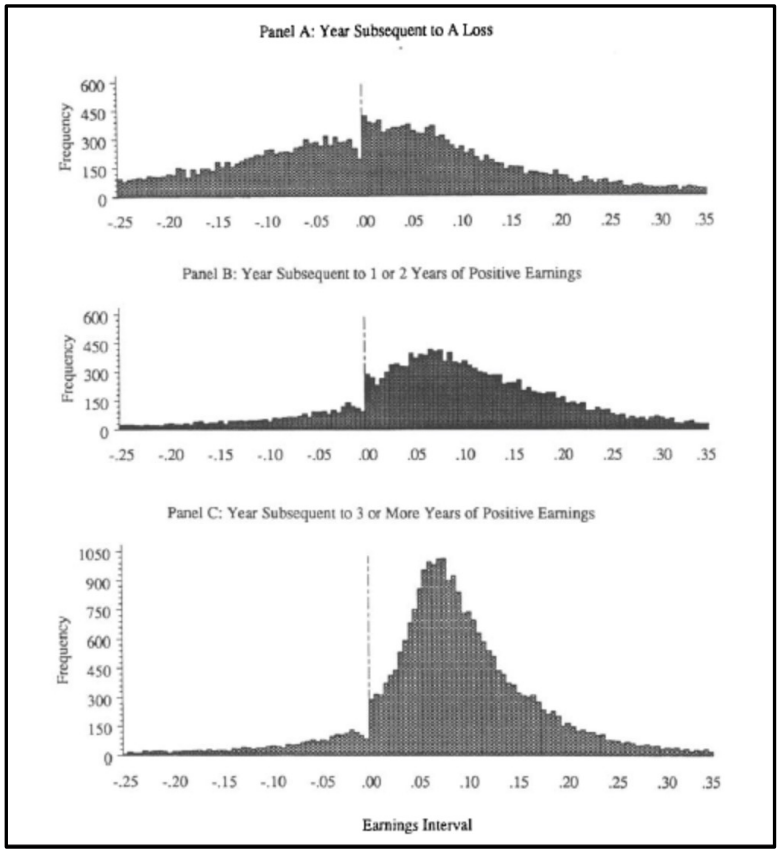



Evidence from 1976-1994, for over 77k firms finds a high significant deviation around zero

Source: Burgstahler and Dichev "Earnings Management to Avoid Earnings Decreases and Losses" (1997), *Journal of Accounting and Economics*

# Earnings Management: Evidence

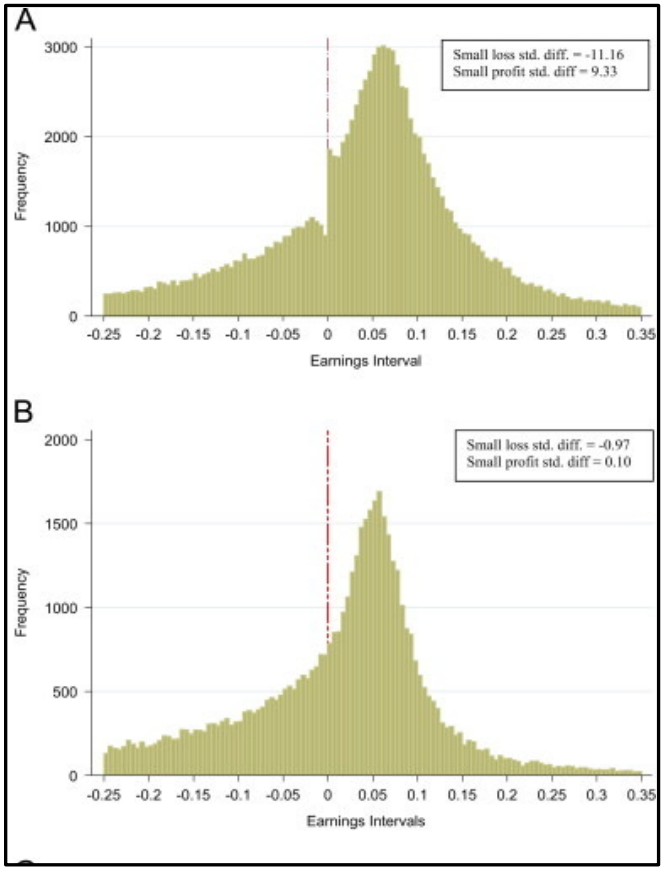
Earnings have a notable discontinuity around zero – that is, firms want to avoid showing an earnings loss




The longer and stronger the trend, particularly positive, the more management doesn't want to show a break …. Or the more they have managed in the past, the more likely they are to be managing today …. Habitual Earnings Managers!

# Earnings Management: Evidence

Does this phenomena still exist today?




Top graph reproduces Burgstahler and Dichev for 1974-2002.
Bottom graph is post 2003-2014. What happened in 2002?
Sarbanes-Oxley Act which carried very stiff penalties for “financial shenanigans”

Source: Gilliam et al “Evidence that the Zero-Discontinuity has Disappeared” (2017), *Journal of Accounting and Economics*

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

1. What does Sloan mean when he says that investors “fixate” on earnings?

   What properties of earnings do investors not understand?

2. What is Sloan’s first testable hypothesis?

3. What is Sloan’s second testable hypothesis?

4. What is Sloan’s main contribution to the literature in this paper? How is it different from other accounting papers?

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

1. What is Sloan's definition of Accruals?

$$\text{Accruals} = (\Delta CA - \Delta Cash) - (\Delta CL - \Delta STD - \Delta TP) - Dep$$

where
- $\Delta CA$ = change in current assets (Compustat item 4),
- $\Delta Cash$ = change in cash/cash equivalents (Compustat item 1),
- $\Delta CL$ = change in current liabilities (Compustat item 5),
- $\Delta STD$ = change in debt included in current liabilities (Compustat item 34),
- $\Delta TP$ = change in income taxes payable (Compustat item 71), and
- $Dep$ = depreciation and amortization expense (Compustat item 14).

2. Did Sloan come up with this measure?
3. What are some of the issues with it?

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

1. What are the the three financial variables that Sloan actually tests?

$$\text{Earnings} = \frac{\text{Income from Continuing Operations}}{\text{Average Total Assets}},$$

$$\text{Accrual Component} = \frac{\text{Accruals}}{\text{Average Total Assets}}, \text{ and}$$

$$\text{Cash Flow Component} = \frac{\text{Income from Continuing Operations} - \text{Accruals}}{\text{Average Total Assets}}.$$

2. What does Sloan say about his choice of the denominator? Is it robust to his choice?

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

1. How does Sloan test H1?

$$\text{Earnings}_{t+1} = \gamma_0 + \gamma_1 \text{Accruals}_t + \gamma_2 \text{Cash Flows}_t + \upsilon_{t+1}$$

2. What does he find?

**TABLE 3**
**Results from Ordinary Least Squares Regressions of Future Earnings Performance on the Accrual and Cash Flow Components of Current Earnings Performance (t-statistics in parentheses)**
**Sample Consists of 40,679 Firm-years from 1962 to 1991[a]**

$Earnings_{t+1} = \gamma_0 + \gamma_1 Accruals_t + \gamma_2 Cash\ Flows_t + \upsilon_{t+1}$

*Panel A: Regressions using actual values*

| | *Pooled* | *Industry Level* | | | |
|---|---|---|---|---|---|
| | | *Mean* | *Q1* | *Median* | *Q3* |
| $\gamma_0$ | 0.011 (24.05)** | 0 019 | 0.009 | 0.016 | 0.022 |
| $\gamma_1$ | 0.765 (186 53)** | 0.721 | 0.635 | 0.703 | 0.780 |
| $\gamma_2$ | 0.855 (304.56)** | 0.781 | 0.733 | 0 777 | 0.873 |

F-test of $\gamma_1 = \gamma_2$: 614.01[b] Proportion of cases in which $\gamma_1 < \gamma_2$: 86%[c]

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

1. How does Sloan test H1?

$$\text{Earnings}_{t+1} = \gamma_0 + \gamma_1 \text{Accruals}_t + \gamma_2 \text{Cash Flows}_t + \upsilon_{t+1}$$

2. What does he find?

**TABLE 3**
**Results from Ordinary Least Squares Regressions of Future Earnings Performance on the Accrual and Cash Flow Components of Current Earnings Performance (t-statistics in parentheses)**
**Sample Consists of 40,679 Firm-years from 1962 to 1991[a]**

$Earnings_{t+1} = \gamma_0 + \gamma_1 Accruals_t + \gamma_2 Cash\ Flows_t + \upsilon_{t+1}$

*Panel A: Regressions using actual values*

| | *Pooled* | *Industry Level* | | | |
|---|---|---|---|---|---|
| | | *Mean* | *Q1* | *Median* | *Q3* |
| $\gamma_0$ | 0.011 (24.05)** | 0 019 | 0.009 | 0.016 | 0.022 |
| $\gamma_1$ | 0.765 (186 53)** | 0.721 | 0.635 | 0.703 | 0.780 |
| $\gamma_2$ | 0.855 (304.56)** | 0.781 | 0.733 | 0 777 | 0.873 |

F-test of $\gamma_1 = \gamma_2$: 614.01[b] Proportion of cases in which $\gamma_1 < \gamma_2$: 86%[c]

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

1. What evidence does Sloan present that investing going long the lowest accruals to cash companies and shorting the highest accruals to cash “makes money”?

   Table 6

   Interpret it for us

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

**TABLE 6**

**Time-series Means of Equal Weighted Portfolio Abnormal Stock Returns Sample Consists of 40,679 Firm-years Between 1962 and 1991[a]**

| *Portfolio Accrual* | *Size Adjusted Returns[b]* | | | *Jensen Alphas[c]* | | |
|---|---|---|---|---|---|---|
| *Ranking* | *year t+1* | *year t+2* | *year t+3* | *year t+1* | *year t+2* | *year t+3* |
| Lowest | 0.049 | 0.016 | 0.007 | 0.039 | 0.007 | 0.001 |
| | (2.65)** | (1.17) | (0.55) | (2.01)* | (0.40) | (0.08) |
| 2 | 0.028 | 0.019 | 0.006 | 0.020 | 0.022 | 0.012 |
| | (3.60)** | (1.65) | (0.68) | (1 68) | (1.53) | (1.06) |
| 3 | 0.024 | 0.012 | –0.006 | 0.018 | 0 014 | –0.006 |
| | (3.84)** | (2.27)* | (–0.86) | (1.70) | (1 28) | (–0.72) |
| 4 | 0.012 | 0.001 | 0.020 | 0.017 | 0.002 | 0.017 |
| | (1.66) | (0.05) | (2.72)* | (2.09)* | (0 17) | (1.29) |
| 5 | 0.001 | 0 002 | 0.006 | 0.010 | 0 004 | 0.014 |
| | (0.03) | (0 22) | (0.86) | (0.87) | (0 38) | (1 12) |
| 6 | 0.010 | 0.005 | 0.016 | 0 006 | 0.002 | 0.003 |
| | (1.43) | (0.72) | (1.90) | (0.57) | (0 24) | (0.43) |
| 7 | –0.002 | 0.003 | –0.006 | 0 004 | 0 006 | 0.005 |
| | (–0.22) | (0.60) | (–0.83) | (0.39) | (0.97) | (0.56) |
| 8 | –0.021 | –0.002 | –0.001 | 0 011 | –0 004 | 0.002 |
| | (–3.03)** | (–0 31) | (–0.01) | (–1.17) | (–0 39) | (0.16) |
| 9 | –0 035 | –0.018 | –0.015 | –0.028 | –0 012 | –0.012 |
| | (–3.70)** | (–2.52)* | (–1.60) | (–3.04)** | (–1 36) | (–1.15) |
| Highest | –0.055 | –0 032 | –0 022 | –0.064 | –0.040 | –0 036 |
| | (–3.98)** | (–2.25)* | (–1.61) | (–4.68)** | (–2 87)** | (–2.47)* |
| Hedge[d] | 0.104 | 0.048 | 0.029 | 0.104 | 0 048 | 0 038 |
| | (4.71)** | (3.15)** | (1.64) | (4 42)** | (2 41)* | (1.62) |

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

1. What evidence does Sloan present that investing going long the lowest accruals to cash companies and shorting the highest accruals to cash “makes money”?

2. How consistent is the return? And do you agree with way for testing consistency?

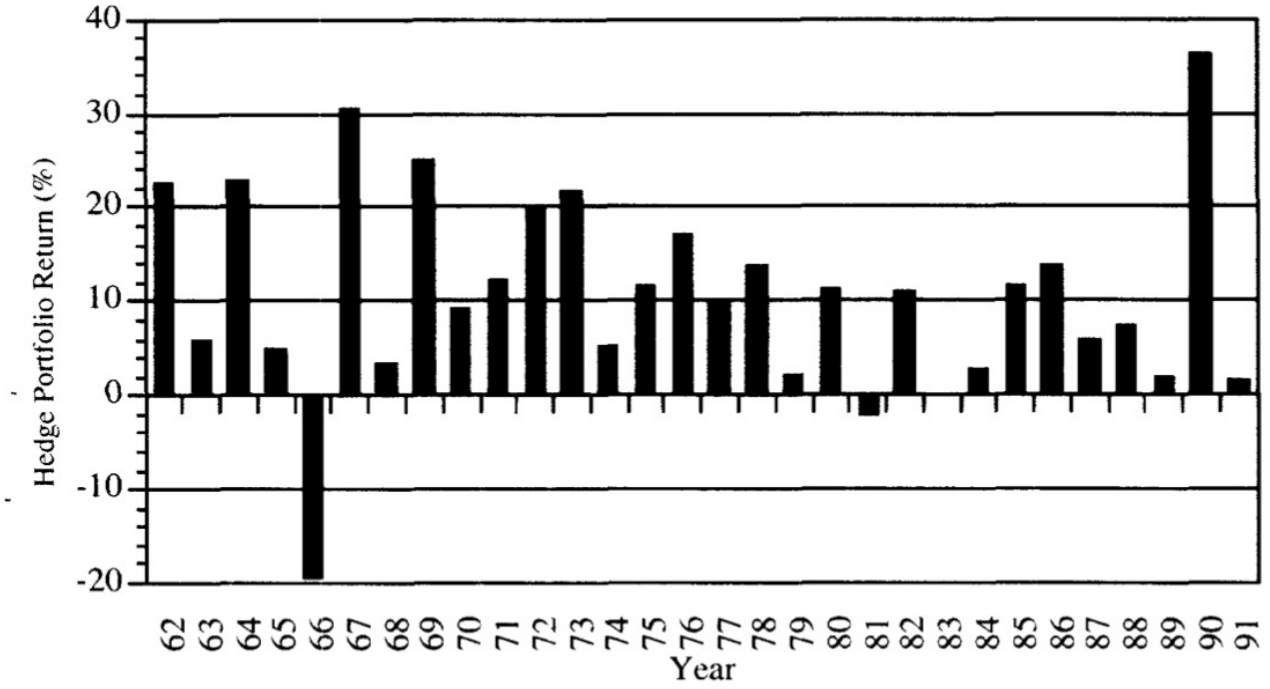
   Figure 2

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

**FIGURE 2**

**Returns by calendar year to a hedge portfolio taking a long position in the stock of firms in the lowest decile of accruals and an equal-sized short position in the stock of firms in the highest decile of accruals. Returns are cumulated over a one-year period beginning four months after the fiscal year end. Accruals is the change in non-cash current assets, less the change in current liabilities (exclusive of short-term debt and taxes payable), less depreciation expense, all divided by average total assets.**



# Earnings Management: Evidence

Sloan (1996) Accrual Paper

1. What evidence does Sloan present that investing going long the lowest accruals to cash companies and shorting the highest accruals to cash “makes money”?

2. How consistent is the return? And do you agree with way for testing consistency?

3. Do accruals help explain the cross-sectional variation in returns and how does he test that?

   Fama-McBeth cross-sectional regressions! Table 7, Panel A

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

**TABLE 7**

**Cross-sectional Regression Tests of the Explanatory Power of Accruals with Respect to Future Annual Stock Returns (t-statistics in parentheses). Sample consists of 40,679 firm-years between 1962 and 1991[a]**

*Panel A: Cross-sectional regressions of stock returns on accruals*

| | *Dependent variable is stock return for:* | | |
|---|---|---|---|
| | *year t+1* | *year t+2* | *year t+3* |
| Intercept | 0.16 | 0.16 | 0.17 |
| | (4.43)** | (4.42)** | (4.29)** |
| Accruals | –0.32 | –0.16 | –0.14 |
| | (–4.60)** | (–2.75)** | (–2 61)* |

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

1. What evidence does Sloan present that investing going long the lowest accruals to cash companies and shorting the highest accruals to cash “makes money”?

2. How consistent is the return? And do you agree with way for testing consistency?

3. Do accruals help explain the cross-sectional variation in returns and how does he test that?

4. Is accruals subsumed by the other known anomalies?

   See Table 7, Panel C

# Earnings Management: Evidence

Sloan (1996) Accrual Paper

*Panel C: Cross-sectional regressions of stock returns on accruals and other predictors of returns*

| | *Dependent variable is stock return for:* | | |
|---|---|---|---|
| | *year t+1* | *year t+2* | *year t+3* |
| Intercept | 0.28 | 0.27 | 0 22 |
| (t-statistic) | (5.60)** | (5.42)** | (4.86)** |
| Accruals | –0 30 | –0.10 | –0 11 |
| (t–statistic) | (–6.15)** | (–2.19)* | (–1.74) |
| Size | –0.02 | –0.02 | –0.01 |
| (t–statistic) | (–3.24)** | (–2.75)** | (–2.08)* |
| Book–to-market | 0.03 | 0.02 | 0.02 |
| (t-statistic) | (2.31)* | (1.97) | (2.32)* |
| Beta | –0 02 | –0.02 | –0.01 |
| (t-statistic) | (–0.91) | (–0.77) | (–0.54) |
| Earnings-to-price | 0.16 | 0.05 | 0.00 |
| (t-statistic) | (2 04)* | (0.55) | (0.01) |

Question: What is the interpretation of the positive intercept?

# Earnings Management: Evidence

Kothari et al (2005)

1. What is the motivation for this paper?
2. Does any one know what the Jones (1991) and modified Jones (1995) model are?
3. What do Kothari et al conclude? What is its impact on Sloan (1996)?

Soliman

1. What does this paper / article tell us? (Hint what does Rothman (2007) multivariate model from Lehman Brothers say about the accrual effect?)
2. Soliman's journey. Anyone know where he worked in "high finance"? Anyone know why he left there?

> It's also the end of a long chapter in Soliman's life that led him from academia to high finance and back again.
>
> "I've come full circle," he says.

# Earnings Management: Evidence

Bongaerts et al (2021)

Ask yourself the same questions we did for Sloan (1996)

Be prepared to discuss!

What is the motivation?

What is the contribution?

What is the key table / result?

What do you think are the strengths and weaknesses? (Hint go back to the very beginning of this lecture …. How many of those questions does the paper address?)

# Natural Language Processing

# Economic Linkages

1. Cohen and Frazzini (2007)

   What is the motivation and central hypothesis of CF?

   What is the contribution?

   What is the key table / result?

   Almost every researcher I know has not been able to replicate the CF result. What part(s) of their methodology might make it very difficult for a research to replicate their result?

   Answer:

   – How do you parse the names of suppliers and the customers?

   – How do you infer appropriate company names?

   – How do you match them to a traded company?

# Challenges in Reading Financial and Accounting Text

Read Loughran and McDonald (2016)

"The meaning of the characters is not unambiguous and in most cases depends substantively on the context of a sentence, document, collection, and when and by whom it was written."

In Finance and Accounting, we are not only looking for the *intended* message but also the *unintended* message.

For example, we might care where a manager unintentionally use weaker words such as "may", "could", "might", "possibly" to potential trouble with the firm.

To do this means we need both a time-series of a firm's language and a cross-sectional measures of firms' language.

# Challenges in Reading Financial and Accounting Text

Where can errors start to creep in?

1. HTML code has not always been consistent. The filing standards, tags and formatting have changed repeatedly over-time for documents such as SEC filings.
   - Tendency to lack structural consistency and anchors is not random
   - Depends upon time periods, regimes, firm size, and industry

2. You need to be highly aware of how certain words will influence your conclusions
   - If you use "best" as a positive word and you don't exclude firm names then *Best Buy* will have unusually positive sentiment scores.
   - If you are capitalization agnostic and use "*may*" then you will have distinct seasonal patterns in sentiment around the month of May.
   - Most common dictionaries use *mine* and *death* as negative sentiment words. But the mining and health industries tend to extraordinarily negative sentiment

All of this leads to issues of replicability because most papers don't give delve into these details.

# Challenges in Reading Financial and Accounting Text

What is meant by readability?

1. Most research has tended to use the FOG Index.

$$\text{Fog Index} = 0.4\,(\text{average number of words per sentence} + \text{percent of complex words})$$

It was developed to determine the number of years of education required to understand the text

Seminal papers in the field use FOG Index:

- Li (2008) on annual report readability;
- Miller (2010) and Lawrence (2013) on retail participation in a stock
- Lehvay et al (2011) on earnings dispersion and forecast accuracy
- DeFranco et al (2013) on trading volume analyst reports and volume

# Challenges in Reading Financial and Accounting Text

## What is meant by readability?

So what's wrong with FOG?

- What is it actually measuring? If you randomly mix the order of the words so the sentence is unintelligible, you get the same FOG score.
- Syllable count is a poor measure of understandability in financial and accounting documents.
  - Loughran and McDonald (2014) find over 50% of FOG scores are coming from words with 2+ syllables
  - Two+ syllables "complex" words include "financial", "management", "operations", "employees", "earnings", "revenues", "customers".
- What do we do with graphics and tables which might actually be improving readability?

What do we do with "jargon"? Is it increasing or decreasing clarity or readability?

You ultimately cannot separate out "complexity" from "readability"

And complexity of the document is ultimately going to be tied to complexity of the business

# Challenges in Reading Financial and Accounting Text

So what do people do?

1. Targeted Phrases are less likely to be amibiguous
2. Word Lists: Henry (2008), Harvard General Inquirer (HGI), Diction, and L&M (2011)
3. Zipf's Law: are the wrong words driving the count? Vice (president); liability; tire; depreciation
4. Term Weighting: not all negative words are equally negative
5. Latent Semantic Analysis, probabilistic LSA and latent Dirichlet allocation: factor analysis for words
6. Document Narrative:
    i. Collocation of words;
    ii. Syntax (grammatical structure);
    iii. Semantics (meaning within a sentence);
    iv. Pragmatics (sentences directly preceding and proceeding the word)
    v. Discourse (topic under discussion)
    vi. Formality of Language Structure

**Guess where is NLP today?**

# Challenges in Reading Financial and Accounting Text

## Similarity

1. Large literature looking at "Similarity of Text"
   i. Similarity Company Documents Cross-Sectionally and Time Series:
      - Annual reports; Products descriptions; MD&A disclosures; Risk Disclosures
   ii. Similarity of Other Text about Company or its Products
      - Analyst reports; Yelp reviews; Amazon reviews; News Articles

2. Almost all measures of similarity use Cosine Similarity:

Given two documents $d_1$ and $d_2$ that have been collapsed into two vectors $x$ and $y$ of word counts, the cosine similarity measure for i=1 to N words is defined as:

$$\text{cosine similarity}(d_1, d_2) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}} \quad (4)$$

# Challenges in Reading Financial and Accounting Text

## What Are Outstanding Issues?

1. What really is a “word”?
2. What really is a sentence? How do you handle different types of punction, hyphens, enumerations, abbreviations, etc.?
3. How do you handle pronouns?
   - “*Mount Everest is 29,029 feet tall. It is in Nepal.*” What does “it” refer to?
   - “*Paul, Jane and Mason were going to Matthew and Ross’s office for lunch today. Unfortunately, they were 20 minutes late.”* Who does “they” refer to?
4. Sarcasm. Jokes. Idioms. “Trader Speak”
5. Negations and Double Negatives.
6. Spoken language contains much more information that is hard to parse?
7. Is there one type of approach that is appropriate for all types of texts (newspapers articles, SMS texts, phone calls, tweets, Bloomberg chats, earnings conference calls, financial reports, blogs, Facebook posts, CNBC tv reports)
8. Brute Force Computation vs Linguistics

# Summary

# Summary

- In a crisis of impacting now illiquid assets, why might we expect to see highly liquid assets' value impacted?
- What is the mechanism by which liquidity crisis might impact the highly liquid assets' value?
- What was the role of leverage in the Quant 2007 crisis? If none of the funds were leveraged would there have been a crisis?
- Why did the crisis subside on August 10$^{th}$ ? What does the large snap-back on August 10$^{th}$ tell you about the crisis, broadly speaking?
- Was the fact that most quants use valuation signals a major source of the crisis?
- What was the role of portfolio construction in the quant crisis? Discuss both a risk and liquidity perspective.
- How unique were the signals that most managers were using? How unique were their returns streams? What is your evidence for both? And clearly explain the difference here.
- If you were running a quant fund in August 2007, what actions were you likely to have taken? What actions should you have taken? How might "agency" problems impacted what steps you would have taken? Would these have been different if you were running a long-short fund, a levered long-short fund, or a long only fund?
- How important in the aftermath of the August 2007 should it be for quantitative managers to look for proprietary signals ("unique alpha")? What might be some of the risks to doing this? Answer this question from both a business perspective (Chief Marketing Officer and Chief Executive Officer) (e.g. how your clients will react) and from the perspective of being the Chief Investment Officer.

# Summary

- What do you believe is the best measure earnings management?
- If you as a researcher are able to detect earnings management in companies, why don't you believe the market is as a whole?
- How might investor clienteles (i.e. stocks that disproportionally owned by one class of investors or another, such as dominated by retail investors vs hedge fund managers) impact the efficacy of Sloans accruals strategy?
- Are there other characteristics of stocks that might make them more likely or less likely to engage in earnings management?
- Why do you think that Sloan's Earnings Accrual Strategy disappeared in and around 2002? How much is publication of the paper in 1996 and how much is implementation of SOX? How might you estimate the different roles? If you believe it was the publication of the paper, why hasn't the publication of other anomalies had such a profound impact?
- When do you think you are most likely to see manager's using "Big Bath" strategies? List 3-5 situations.
- How important do you believe it is that Sloan doesn't account for transaction costs in his strategy? Do you believe the entire effect would go away if he accounted for t-costs? If not, why not? Give clear direct and specific evidence (not just opinions as we all know it is better to account for t-costs).
- If you were a member of the FASB (the accounting standards board) and you read Sloan (1996), what if any changes might you want to make to accounting standards (e.g. how firms report earnings or how earnings are calculated)?

# Summary

- You work at an Equities Hedge fund. The strategy has two components (i) Country-Industry selection; and (ii) stock selection within Country-Industry pairs. In other words, one component of the model decides how much to over/under-weight U.S. telecoms relative U.S. technology relative to Japanese Financials relative European Materials relative to Chinese Media companies, etc., etc. The other component of the model decides how to overweight and underweight all the stocks within those countries-industry buckets. Your CIO is very disappointed in the performance of the Country-Industry model for despite the performance still be positive and relatively consistent as it less than 1/3rd the Sharpe Ratio of the Stock Selection model. As the Head of the Country-Industry Model team, how might you respond to your CIOs criticism of your relatively lower SR? (note: no whining allowed that it is hard! If it were easy he wouldn't be paying you seven figures. You need to marshal some quantitative evidence about why it is hard).
- Do you think quantitative asset managers are more or less likely to experience another August 2007 event today?
- Many of the seminal papers in the NLP literature are hard to replicate. What might be some of the reasons for that?
- Many researchers use the concept of a text's readability as a signal in their models, with the idea being that the less readable the text is the more likely it is that management is trying to hide information from investors and analysts. What are the some of the ways researchers try to measure readability and what their flaws and strengths?
- You notice that in your model of readability that the sector called "conglomerates" persistently ranks at the bottom of readability scores. Why is this likely to be the case?

# Summary

- Why is the tone of a financial earnings report or transcript a very difficult construct (concept) to measure?
- You are a regulator trying to make financial documents more readable to the average reader. You proposed a pilot study 18 months ago to a series of firms for them to follow. You know have 18 months of their data using your pilot study guidelines along with the traditional reports that they also had to follow. How would measure if your pilot study disclosure forms were more readable than the pro forma financial disclosure forms?
- What is meant by the term “linkages” across firms? Why might they be important and how might you measure them?
- Your researcher has come to you and said that the parsing the text directly and trying to understand its meaning is a hopeless cause. Instead, they want to look for “touples” of text that either is strings of words separated by a “ “(space) or a string of characters followed by one of the following symbols: “+” “-” “;” “:” “/” “!” “?” “.” “=“ “(“ “)” “<“ “>”. They then want to take headlines of stories and the entire set of “touples” in them and test when a story is published on Bloomberg (which has a point in time stamp that is ultra accurate and reliable) and determine what is the next 60 min return of the stock. For purposes of this question assuming that there are no revisions to a story and it is only published once.
  - Why does your researcher think that parsing the text directly is a hopeless cause? Do you agree or disagree with them?
  - What do you think of their strategy of testing “touples” with returns directly? Strengths and weaknesses?
  - Do you agree with all the “touples” they are suggesting? Why or why not? Give specific examples where they might make sense and where the association with returns is less obvious?
  - Do you think this approach makes sense to use for languages other than English (with appropriate modifications for characters and punctutation)? Would you trust the results?