EFR summary XX

Introduction to Accounting Research,

FEM11110

2025-2026

Weeks 1 to 6

Details

Subject: Introduction to Accounting Research 2025-2026

Teacher: Dr. FM Elfers

Date of publication: 12.10.2025

Koffietje doen?

Start jouw carrière bij BDO

Maak kennis met BDO Accountants & Adviseurs, de beste plek om als toptalent aan de slag te gaan. De koffie staat voor je klaar. Vertellen wij je over wat jij kan bijdragen, en jij ons over je ambities.

Scan de QR-code en plan jouw koffiemoment in.

werkenbijbdo.nl ▶

BDO

Introduction to Accounting Research

Lecture 1

Accounting research refers to the use of quantitative (primarily financial) and qualitative information by economic actors. Accounting research is an applied form of research that is relatedto its adjacent fields but with a specific focus on information. IASB and FASB often use academic input for standard setting decisions. Relevant insights, also for investors (e.g. trading opportunities).

Areas of accounting research

Financial accounting (FA)
• Deals with how managers produce financial information for economic agents (investors, financial analysts, etc.) outside the organization and how these agents respond to variations in accounting methods and estimates.

Auditing (as a sub-discipline of financial accounting)
• Deals with the general question how auditors audit financial statements and examines the antecedents and consequences of variations in auditor characteristics and work methods.

Managerial accounting (MA)
• Examines how economic agents within the organization (managers and employees) produce and use accounting information.

Accounting research is based on theoretical models, interested in causal relationships (focus on “endogeneity concerns” and ”internal validity”) and applying frequentist inferential statistics (e.g. obsession with p-values). Is accounting research “Data Science”? → Yes and no. Published empirical accounting research mostly represents only a subset of data science problems

Research questions
A good research question addresses the relation between two concepts/constructs (X & Y), can be stated clearly and unambiguously as a question, implies the possibility of empirical testing, and is important to the researcher and others.

How to identify good research questions, e.g., for a thesis?
→ Talk to practitioners, popular press (identify current issues), identify topics that standard setters (IASB), regulators (AFM, SEC), consultants and other institutions (CFA, CPA) are currently discussing, identify unexplored issues in the academic literature, think about whether the average results found in prior research might be stronger or weaker for particular firms, or in particular settings.

How to structure an (empirical) research project: Formulate research question → Literature review → Theory development → Formulate hypotheses → Empirical analysis → Conclusion and implications

Lecture 2

Purpose of theoretical/analytical studies: provide deeper insights and develop causal predictions

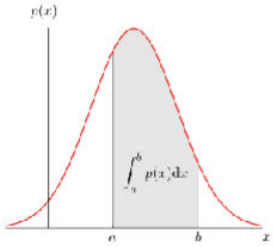
DeAngelo (1981)
• Low balling: asking a lower price (discount) in the first period, so you can raise the price in the future.
• If you have got a client, you can get future quasi rents (future audit fees – current audit fees) because of the investment and switching costs.
• Competition drives profits to zero
• The expectation of future rents reduces independence because these rents are only collected if the client does not terminate the engagement.
• According to DeAngelo, the expected quasi rents results in a lower auditor independence and cause

the association between low balling and auditor independence.

Questions on DeAngelo
• A price floor for initial audits will not increase auditor independence, auditors would find different tools, maybe they will invite them e.g. for a diner. They will find a way to compete and that will be costly again.
• A other solution that would increase auditor independence in this model is a mandatory audit rotation, reducing the amount of future quasi rents. Or increase fines for auditor misbehavior.

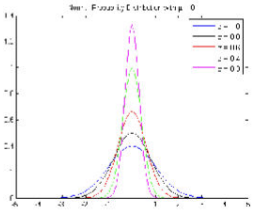
Empirical research/statistics refresher

Random variable: assigns a unique value to the outcome of a random experiment, associates exactly one of the possible outcomes to each trial of a random experiment. (e.g. randomly picking one person from a group and measure body height).
Categorical (qualitative) variables: take category or label values and places a unit of observation into one of several groups (e.g. nationality, hair color, etc.)
Quantitative variables: discrete random variable, a random variable that has a finite set of specific possible values (e.g. throwing a dice).
Continuous random variable: can take on a continuous range of possible values.



Cumulative distribution function: indicates the probability of a random variable taking values below or equal to a certain value. Alternative representation: **probability density function** (indicates the relative likelihood of a random variable to take a specific value. →
• The area under the PDF between two values is the probability that an observed value will fall between these thresholds
• The total area under the PDF is 1

In the easiest case of a **normal distribution**, the distribution can be characterized only by the mean (μ) and the variance/standard deviation of the random variable. Standard deviation σ is square root of variance σ^2.
In a normal distribution
• 90% of all observations are within 1.645 σ around μ
• 95% within 1.96 σ around μ
• 99% within 2.576 σ around μ



Population: all individuals/companies/subjects that are of interest to a researcher.
Sample: a subset of the population

Sample selection bias: bias resulting from non-random sampling of observations from the population

Statistics calculated from randomly picked samples represent a random variable as well, and they also have a distribution, called the "sampling distribution".

Standard deviation of all sample means is $\frac{\sigma}{\sqrt{n}}$

So, if we want to infer the true population mean μ of a variable X from a sample:
• We can calculate the mean $\bar{x}$ from a random sample
• We know that $\bar{x}$ is an unbiased estimator of μ (because the mean of sample means $\bar{x}$ is μ)
• We know that the distribution of $\bar{x}$ around μ is normal with a standard deviation of $\frac{\sigma}{\sqrt{n}}$

$\bar{x}$ is a point estimate of μ, and we can say something about how accurate this point estimate is by:
Confidence interval indicates how sure we are that the true value μ will be within a certain range of values around $\bar{x}$: $\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}$ (multiplier z depends on the desired level of confidence)

In some cases, when we don't know the population mean, we can replace σ by s (the sample standard deviation) and the confidence interval becomes $\bar{x} \pm t * \frac{s}{\sqrt{n}}$, where $\frac{s}{\sqrt{n}}$ is the standard error of $\bar{x}$

In terms of standard error, the sample mean is not normally distributed, but follows a t-distribution that depends on the sample size (with n – 1 degrees of freedom).

Hypothesis test: We reject a H_0 if the p-value is lower than a 'significance level' *a* of 0.1, 0.05 or 0.01

In case of directional hypotheses, you can also use a one-sided test. (e.g., H1: "... is larger than zero")

Summary:
1) Draw a sample of *n* observations from the population
2) Calculate the sample mean $\bar{x}$
3) Calculate the standard error of the sample mean $\bar{x}$: $\frac{s}{\sqrt{n}}$
4) Calculate a test statistic
5) If the t-stat is larger (or smaller) than the critical values of the corresponding t-distribution with n – 1 degrees of freedom, reject the null hypothesis.

For hypothesis testing, two types of errors may occur:
• Type 1 error: rejecting a null hypothesis that is actually true
• Type 2 error: accepting a null hypothesis that is actually false

'Size' of a test: probability of falsely rejecting the null hypothesis when it is actually correct (type1error)
'Power' of a test: probability of rejecting the null hypothesis when it is indeed false (1 – probability of Type 2 error)

In general, the power increases when the effect size is larger and the sample size is larger
→ In accounting research, most people focus on significance only.

It is important to also think about the "economic significance" – does the effect really matter?

General problems in economics (and accounting): strong publication bias for significant results.
• "p-hacking": play around with specifications until you get a result with a p-value below the arbitrary benchmark of 0.1 or 0.05.
• "HARKing"; (Hypothesizing after the results are known): ex post nearly every result can be "plausibly" explained.
• "Outcome switching": Check various outcome variables and report only those with significant results.

The predictive validity framework

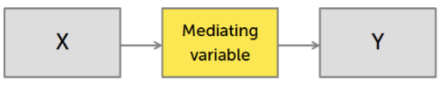
Construct: Abstract idea which is not directly observable or measurable and should be operationalized for empirical testing. From the theory domain. (e.g. intelligence, firm performance)
• "Construct" and "concept" are often used interchangeably.

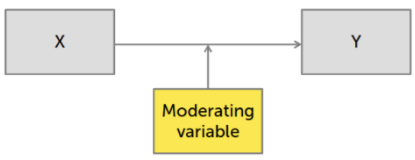
Variable: Observable item which can assume different values and is used to measure a theoretical construct. Used in the empirical analysis. (e.g. IQ, ROA, CO2 emissions)

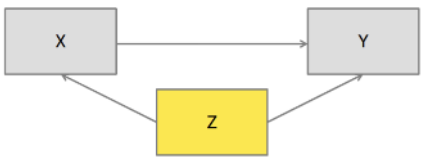
Important determinants of the validity of the empirical analysis:
• Construct validity: The degree to which a measurement/variable ("operationalization" of a construct) captures the underlying theoretical construct it is supposed to measure. E.g. to what extent does IQ capture the construct of intelligence?
• Reliability: The degree to which a measurement provides estimates which are consistent. E.g. if you measure the same person's IQ 10 times, do you get 10 similar results?

Each of these (causal) questions relates to the effect of X on Y.
• X = Independent or Explanatory or Right Hand variable
• Y = Dependent or Explained or Left Hand variable

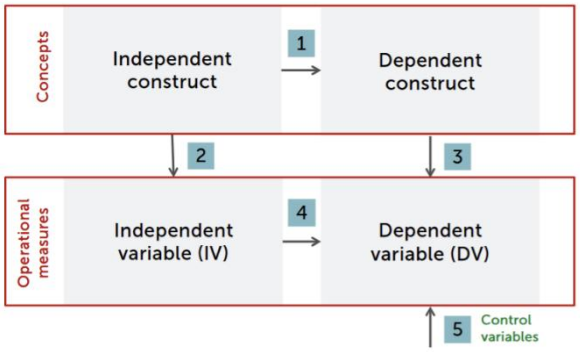
In applied research, you will encounter additional terminology to classify variables.
• Mediating ("intervening") variable: A variable that explains the mechanism between the independent and dependent variable by splitting the relation between them in two parts.
• Moderating variable: Factors that strengthen or weaken the relation between a dependent variable







and an independent variable.
• Control variables: capture the effects of other factors (Z) that are related to Y or both X and Y, but which are not of direct interest to us when examining the effect of X on Y.



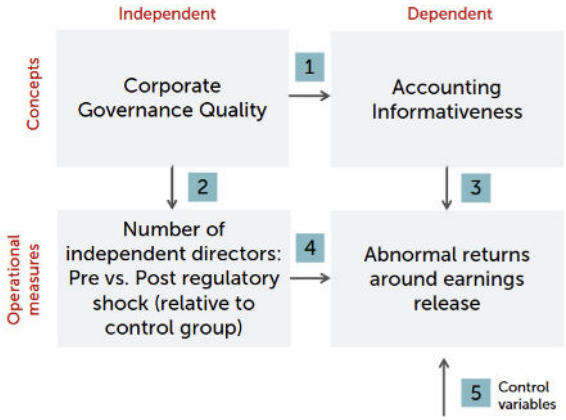
The **Predictive Validity Framework from Libby** (1981) is extremelyhelpful in setting up a research study and its research design. Also known as "Libby boxes". 4 boxes, 5 links.

Link #1 captures the hypothesized causal relation. The boxes reflect the theory domain of the concepts of interest.

Links #2 and #3 reflect operationalizations/measurements of Xand Y. Construct validity and reliability play a crucial role here

Link #4 is the (causal) relation we are empirically testing.

Link #5 reflects the effect of other factors (control variables) on the outcome Y.

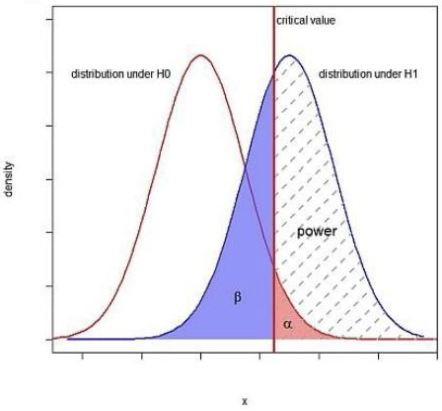


Question: Have a look at this hypothetical abstract of a paper: "Agency theory predicts that firms with a higher quality of corporategovernance produce more informative accounting numbers. Weexamine a regulatory shock that substantially increased theproportion of independent directors in one set of firms but not inother firms, and find that the stock market reaction to earningsreleases (measured by abnormal returns) increased after this changein regulation, but only for those firms affected by the regulation."

Can you create (and label) the "Libby Boxes" for this paper? →

Lecture 3

Recap hypothesis testing:
• Develop a hypothesis (and null hypothesis)
• Draw a random sample of observations from the population
• Calculate a statistic from the sample as an estimate of the population parameter
• Generate a test statistic that reflects the likelihood of obtaining the observed statistic assuming the null hypothesis were true.
• Compare p-value associated with the test statistic to the required significance level a

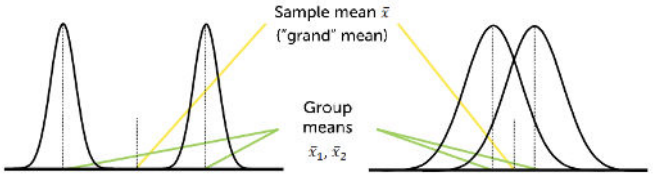


Sampling distributions ⟶
a = likelihood of a type 1 error, β = likelihood of a type 2 error

T-tests and ANOVA

Relationships between two variables: how does one variable influence another? If you only have two groups, you can use two-sample t-test.

ANOVA is a generalization of the two-sample t-test. For $I = 2$ groups, these two tests are equivalent, but ANOVA can be used for any number of groups. ANOVA tests whether the observed differences in means of a variable X across I groups within a sample are statistically significant. For example:
• H_0: In the population, the groups don't matter and therefore all group means are the same
• H_1: At leas one of the groups has a different mean than the others.



ANOVA can cover a special case when the independent variable is qualitative/categorical

Idea: decompose total variance of a sample in a between-group and within-group component ("error").

Illutstration: one "factor" (the independent variable) that divides the sample of n observations into $I = 2$ groups. (e.g. education: yes/no bachelor degree). Look at the distribution of observed values of X across groups (e.g. salary, do people with and without a BSc earn different salaries?)

• Left, a big *between variation*, and a small *within variation*.

- Total Sum of Squares ("SST"): $SST = \sum_i \sum_k (x_{ik} - \bar{x})^2 = SSA + SSE$
- Between groups: $SSA = \sum_i n_i (\bar{x}_i - \bar{x})^2$
  - This is the SS you would get if all observations in one group had the same value (the group average for groups according to factor A, hence "SSA"), i.e., the variation "explained" by the factor.
- Within groups: $SSE = \sum_i \sum_k (x_{ik} - \bar{x}_i)^2$
  - This is the SS of the difference between each observation k in its group i and the respective group mean (the "error", or "residual", hence "SSE").

• Right, a small *between variation*, and a big *within variation*.

- i: group index
- k: individual index
- n_i: number of observations in a group
- $\bar{x}_i$: group mean
- $\bar{x}$: sample mean

If *between variation* > *within variation*, then it seems unlikely that in the population there are no differences between groups.
If *between variation* > *within variation*, it is more likely that there are no differences in the groups

But, what is a 'large' difference? → compare the mean squares for between and within variation
Mean squares: sum of squares (SS) of deviation of observed values from predicted values divided by respective degrees of freedom.

Dividing both terms by their respective degrees of freedom yields the mean squares.

The ratio of the mean squares (MSA/MSE) gives you the ANOVA test statistic called the F-statistic, as it follows a F-distribution with (d1,d2) degrees of freedom.
• A larger F-statistic means that the between variation is large relative to the within variation, and therefore provides evidence against H_0)

ANOVA can fit many factors as long as they are categorical. If you add an additional factor B with J different groups, this gives you $I * J$ cells. Total sum of squares now: $TSS = SSA + SSB + SSAB + SSE$
• $SSAB$ is calculated from the difference between each cell mean and what you would predict as the cell mean given the combination of the other factors ('interaction effect').
• SSE is calculated from the difference between each single observation k and its respective cell mean

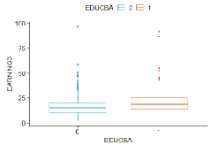
The F-statistic for the significance tests is calculated as the ratio of the respective mean squares and the mean squares of the error.

Example:

Does a person's future income depend on their education?

- First just one factor: bachelor degree yes/no (EDUCBA = {0,1})
  - Dependent variable: hourly salary (EARNINGS)
- H_0: Obtaining a bachelor degree does not affect a person's future income.
- Data: n=500 observations from a longitudinal study in the US (1979-2002)[1]
- Descriptive statistics:

```
> by(dat$EARNINGS, dat$EDUCBA, summary)
dat$EDUCBA: 0
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2.00   10.73   14.75   16.75   19.90   96.15
------------------------------------------------
dat$EDUCBA: 1
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   4.29   13.50   18.75   21.14   25.00   91.67
>
```



- R output for ANOVA:

```
> # Run ANOVA
> res.aov <- aov(EARNINGS ~ EDUCBA, data = dat)
> # Show results
> summary(res.aov)
             Df Sum Sq Mean Sq F value   Pr(>F)
EDUCBA        1   2073  2072.6   17.76 2.98e-05 ***
Residuals   498  58128   116.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

- F = Mean square of treatment / Mean square of residual = 2072.6 / 116.7 = 17.76
- F-distribution with (d1,d2) degrees of freedom, where:
  - d1 = 1 (number of groups minus one)
  - d2 = n-2 = 498 (number of observations minus number of groups)
  - Critical value F(1, 498) = 3.87 at 5% significance level
  - Critical value F(1, 498) = 6.69 at 1% significance level

Paper Van Rinsum, Maas, Stolker (2018)

• Topic: Auditing. Method: laboratory experiment with experienced auditors
• ANOVA is rarely used in archival empirical research. However, it is a common approach in experimental research, where there is a treatment and control group.

RQ: Do decision aids in auditing (e.g. checklists) affect the quality of auditor judgment an decision making on related, but distinct, tasks?

Theory:
• Auditors are more likely to accept aggressive accounting when using checklists.
• Auditors are more likely to be pressured by management than by independent audit committee.
• Pro-client bias will be higher when auditors can retain the illusion of objectivity by referring to satisfactory checklists.

Hyptoheses:
1. Use of disclosure checklists increases perceived acceptability of aggressive accounting treatments
2. This effect is more pronounced when the auditor is appointed by management than when the auditor is appointed by an independent audit committee.

Operationalization:
• Experiment with 55 experienced auditors
• Independent variables: 1 group with a disclosure checklist, 1 group without
• Dependent variable: 2 audit cases + an exit questionnaire

Main effect: SLIDE 32
Interaction/moderating effect: SLIDE 32

Regression basics (ordinary least squares)

ANOVA is useful when we are interested in differences between groups. But in many cases, the independent variable is a continuous variable, and ANOVA does not work here. Instead use OLS.

Regression: How does Y respond to a one unit change in X?
• Conditional expectation function: $Y = E(Y|X) + \varepsilon$
• Assuming a linear relationship: $Y = a + \beta X + \varepsilon$
  - a is the intercept
  - β is the slope coefficient, capturing the effect of X ('beta')
  - ε is the error term

OLS estimates the parameters in such a way that the sum of the squared value of the n (= number of data points) residuals e is minimized/

Univariate: regression where we only have one independent variable X.
Multivariate: includes multiple explanatory variables

Gauss-Markov assumptions
1. We have a random sample of observations of Y and X from the population

- Now add a second factor and also test the interaction effect: gender (MALE = {0,1})
- H_0: The effect of obtaining a bachelor degree on future income is not different between men and women.
- Descriptive statistics:

```
> library(plyr)
> means.by.factor <- ddply(dat, c("MALE", "EDUCBA"), summarise,
+                mean.EARNINGS=mean(EARNINGS))
>
> means.by.factor
  MALE EDUCBA mean.EARNINGS
1    0      0      16.48988
2    0      1      18.34975
3    1      0      16.99914
4    1      1      24.10355
>
```

- Increase from EDUCBA for men: 7.10
- Increase from EDUCBA for women: 1.86
- But is the difference (5.24) significant?

- Two main effects:
  - EDUCBA (significant at 1%)
  - MALE (significant at 1%)
- And one interaction effect:
  - EDUCBA*MALE (significant at 5%: F = 6.46, p = 0.0113)
  - We can reject the null hypothesis that the effect of EDUCBA is the same among the two groups at a significance level of 5%.

```
> Anova(lm(EARNINGS ~ EDUCBA.f * MALE.f, data=dat, contrasts=list(EDUCBA.
f=contr.sum, MALE.f=contr.sum)), type=3)
Anova Table (Type III tests)

Response: EARNINGS
              Sum Sq  Df   F value    Pr(>F)
(Intercept)   155167   1 1354.8037 < 2.2e-16 ***
EDUCBA.f        2162   1   18.8773 1.692e-05 ***
MALE.f          1055   1    9.2147  0.002527 **
EDUCBA.f:MALE.f  740   1    6.4614  0.011328 *
Residuals      56807 496
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

2. There is indeed a linear relationship between X and Y.
3. No perfect linear relationship among the explanatory variables, i.e. no perfect (multi-)collinearity
4. The conditional mean of the error term is zero: $E[\varepsilon|X] = 0$. That means the error term is 'exogeneous' and does not correlate with any of the explanatory variables.
5. The error terms are 'independent and identically distributed' (i.i.d.) with a mean of zero
   - That means they have the same mean of zero and the same variance (homoscedasticity).

$$\frac{(\hat{\beta}-\beta_0)}{SE_{\hat{\beta}}}$$

$$SE_{\hat{\beta}} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

Homoskedasticity means that the conditional variance of the regression error term is the same for all values of X: $E(\epsilon_i^2|X) = \sigma^2$

$$SE_{\widehat{\beta_h}} = \sqrt{\frac{\hat{\sigma}^2}{(1-R_h^2)\sum_{i=1}^{n}(x_{ih}-\overline{x_h})^2}}$$

Heteroskedasticity means that the conditional variance of the regression error term is not the same for all values of X: $E(\epsilon_i^2|X) = \sigma_i^2 \neq \sigma^2$

Standard error $\hat{\beta}$ is an estimate of all the standard deviation of the underlying sampling distribution.

```
Regression Results
==========================================
                      Dependent variable:
                    ---------------------
                      Annual_Stock_Return
------------------------------------------
ROE                          0.121**
                             (0.060)

Constant                     1.910*
                             (1.099)

------------------------------------------
Observations                   452
R2                            0.009
Adjusted R2                   0.007
Residual Std. Error      19.865 (df = 450)
F Statistic           4.040** (df = 1; 450)
==========================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

- Test statistic for the significance of $\hat{\beta}$ is the t-statistic: divide the difference between the estimated regression coefficient and the benchmark value under the null hypothesis by the standard error.
- Then compare this ratio to the t-distribution's critical values (e.g. 1.65, 1.96 or 2.59)
- In practice, the default benchmark is nearly always zero.

$$\frac{(\hat{\beta}-\beta_0)}{SE_{\hat{\beta}}} = \frac{0.121}{0.06} = 2.01\ (> 1.96)$$

The standard error $SE_{\hat{\beta}}$ for a univariate regression ($Y = a + \beta X + \varepsilon$) can be calculated as

- The coefficients get more precise (smaller SE) when you have more observations, more variation in X and less noise from the error.

For a multivariate regression with H different regressors ($Y = a + \beta_1 X_1 + \cdots + \beta_H X_H + \varepsilon$), the formula for the standard error of a specific regression coefficient is:

- The coefficients get less precise with a stronger linear relationship between the different regressors.
- This is the problem of multicollinearity: your standard errors can become very large, so not significant

How do you calculate R^2?

TSS = total sum of squares = $(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \ldots + (y_n - \bar{y})^2$

RSS = residual sum of squares = $(e_1)^2 + (e_2)^2 + \ldots + (e_n)^2$

$R^2 = 1 - [RSS/TSS]$

Example:

$\hat{\beta} = 0.121$

A one percent point increase in ROE is associated with a 0.12 percentage point increase in annual stock returns.

Is the coefficient significantly different from zero?

- Standard error of $\hat{\beta}$ is 0.060
- So t-stat for $\hat{\beta}$ is

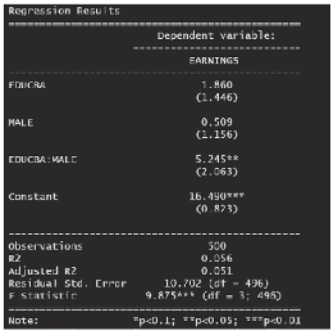
So, we reject the null hypothesis that $\beta = 0$ at a significance level of 5%

R squared (R^2) provides information on the portion of the variation in the dependent variable Y that is explained by the independent variable X, so how 'good' is the regression model in explaining Y?

- The higher the R^2, the better the 'fit' of the model. Ideal value = 1
- A low R^2 does not mean you have a bad model, but there are a lot of other influences on the dependent variable
- Adjusted R^2 equals R^2 with an adjustment for the number of X variables

Interaction effects of two variables X_1 and X_2 are captured by adding their product $X_1 * X_2$ to the regression equation, in addition to their 'main effects' as separate variables.
$Y = a + \beta_1 X_1 + \beta_2 X_2 + \beta_3 * X_1 * X_2 + \varepsilon)$

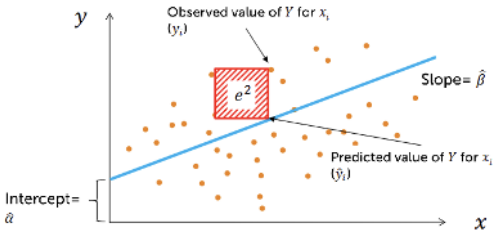
The coefficient on $EDUCBA * MALE$ ($\hat{\beta}_3$) shows the incremental relation for men: their overall effect of EDUCBA is $\hat{\beta}_1 + \hat{\beta}_3 = 1.86 + 5.24 = 7.10$: For men with a BA, salary is higher by 7.10 compared to men without a BA.



Regression Results

	Dependent variable:
	EARNINGS
EDUCBA	1.860
	(1.446)
MALE	0.509
	(1.156)
EDUCBA:MALE	5.245**
	(2.063)
Constant	16.490***
	(0.823)
Observations	500
R2	0.056
Adjusted R2	0.051
Residual Std. Error	10.702 (df = 496)
F Statistic	9.875*** (df = 3; 496)
Note:	*p<0.1; **p<0.05; ***p<0.01

- $\hat{\beta}_0 = 16.49$: Women who do not have a BA degree earn 16.49, on average.
- $\hat{\beta}_2 = 0.51$: Men without a BA degree earn 16.49 + 0.51 = 17.00, on average.
- The coefficient on $EDUCBA$ ($\hat{\beta}_1$) shows the relation for women: salary is higher by 1.86 for women with a BA compared to women without a BA.



Lecture 4

ANOVA (recap)
• Compare means of a variable across groups derived from one or many factors
• Idea: Decompose the total variance of a sample in a between-group and a within-group component
• Calculate 'mean squares': sum of squares of deviation of observed values from predicted values divided by respective degrees of freedom
• Ratio of 'between' mean squares to 'within' (error) mean squared gives you the F-statistic that can be used to infer if differences are statistically significant or not
• If F-statistic is significant, you know that at least one group has a mean that is so different from the other croup that it can hardly be explained by random sampling from the same group

OLS (recap)
• Independent variable X and the outcome variable Y.
• OLS fits an regression line, that represents the relationship between X and Y, with slope β.
• The residual is the distance between the line and the observation (e^2).

OLS regression – Coefficient standard errors
• The validity of the formulas for the standard errors depends on OLS assumptions IV and V
• A violation of these assumptions is a problem, because it generally leads to understated estimates of the standard errors. That means that you are more likely to find wrongly 'significant' results.
• Endogeneity (IV) is a problem
• Violations of assumption V (i.e. heteroskedasticity) are commonplace in practice. You can fix it by using 'robust' or 'clustered' standard errors.

Example: Assume you want to investigate whether the mandatory introduction of certain environmental disclosures at the factory level leads to better environmental performance. However, you believe that such an effect will depend on the level of attention to these disclosures.

You measure:
• Environmental performance as WASTE, the amount of toxic waste produced by a factory.
• Mandatory disclosures as MANDATORY, a dummy variable that is equal to one in all years after the introduction of mandatory disclosure requirements, and zero otherwise.
• Attention as ATTENTION, which you measure by the number of local newspapers active in the county of the factory.

Q: Please come up with a regression model that will capture your research question:

$$WASTE = a + \beta_1 * MANDATORY + \beta_2 * ATTENTION + \beta_3 * MANDATORY * ATTENTION + \varepsilon$$

Nonlinear relationships between variables

OLS is commonly used because of its simplicity. However, OLS is not always applicable.
→ E.g.: the linearity assumption: OLS assumes that there is a linear relation between the independent and the dependent variable, i.e., a one unit increase in the independent variable will always lead to a constant increase/decrease in the dependent variable. Is this realistic?

Non-linear functions cannot be reasonably estimated using basic OLS. But in some cases, you can stick to OLS, but transform the dependent or independent variables, or add additional terms as polynomials. Interaction models (see last week) are also a way to capture nonlinearities.

Variable transformations – Logs

Log transformations, (natural) logarithms: $y = e^x \leftrightarrow \ln(y) = \ln(e^x) = x$
• Multiplicative relationships can be linearized: $\ln(xy) = \ln(x) = \ln(y)$
• Differences between logs of values are approximately equal to the relative difference between the

raw values: $\ln(x_1) - \ln(x_0) \approx \frac{x_1 - x_0}{x_0} = \frac{\Delta x}{x}$

• By applying logs to variables, you 'compress' large values and 'spread' small values

Benefits to use log transformations in regressions:
• Theory driven (estimate multiplicative relationships or de-/increasing marginal effects using OLS
• Variables with a skewed distribution (especially for variables that are only positive (money, age).
• Likely to have results driven by influential outlier observations.
• Can induce heteroskedasticity and non-normal errors.
• Log transformed variables 'pull' the extreme values to the center and reduces these effects

How to interpret regression coefficients based on log-transformed variables?

I. Log-log models (sometimes also called "log-linear"): Both the dependent and independent variable have been log- transformed: $\ln(Y) = a + \beta \ln(X) + \varepsilon$
• If X changes by one percent, then Y changes by about β percent

II. Lin-log models: only the independent variable has been log-transformed : $Y = a + \beta \ln(X) + \varepsilon$
• If X changes by one percent, then Y changes by about $0.01 * \beta$ units

III. Log-lin models: only the dependent variable has been log-transformed : $\ln(Y) = a + \beta X + \varepsilon$
• If X changes by one percent, then Y changes by about $100 * \beta$ units

Variable Transformations - Polynomials

• Another approach is to add quadratic (cubic, quartic,…) terms of X to the regression as polynomials.
• In practice, most of the time you only see the quadratic term: $Y = a + \beta_1 X + \beta_2 X^2 + \varepsilon$
• Usually, β_1 would give you the effect of a unit change in X, holding everything else constant.

But obviously, you can't change X without changing X2: β_1 has no standalone meaning any longer
Instead, the marginal effect of X on Y is defined by $\frac{dY}{dX} = \beta_1 + 2\beta_2 X$ and depends on the level of X
See interaction effects last week: a quadratic term is just an interaction of X with itself.

• Complex polynomials can give a very good fit within you sample, but danger of overfitting
• Also, they are very sensitive to small changes in the dataset

Some alternative regression models

Another problem: OLS is designed for continuous outcome variables. But think of the following questions:
• Binary outcome: Evaluate the determinants of accounting restatements - either yes (1) or no (0): Probit/Logit regression.
• Unordered multinomial outcome: Influence of social background on choice of field of study - sociology, law, medicine, accounting...: Multinomial Logit/Probit.
• Ordered multinomial outcome: Influence of country-level determinants on scope of IFRS adoption - no adoption (0), partial adoption (1), full adoption (2): Ordered Logit/Probit.
• Censored data: Influence of education on income, but survey data is truncated ("more than 100.000 EUR"): Tobit regression.
• Duration: Determinants of bankruptcy, observe firms during a crisis: If they go bankrupt - after how many months. Else - we cannot observe when and if they ever go bankrupt: Hazard/Duration regression.

Binary dependent variables – Logistic Regression

Binary outcomes are commonplace in accounting research. In these cases, you are interested in the probability of an event to take place.

The logistic ("logit") regression is a "generalized linear model" that fits an s-shaped logistic curve to the data. "Probit" regression is a variation of this concept, mostly no difference in large samples.

"**Odds**": probability of something to happen vs. probability of something not to happen.
• So if something happens with probability 0.1 (10%), then the odds are 0.1/0.9 (=0.11) or one to nine"

There is no "closed form" (i.e., unambiguous analytical) solution to estimating the logit parameters.
Instead, based on the iterative "maximum likelihood" methodology: trying out different values until fit can't be improved

Table 3
Logit estimates on bank distress and their predictive performance.

Estimates	(1) Benchmark
Bank-specific indicators	
C[a]	
Intercept	3.46***
Capital ratio	−0.76***
Tier 1 ratio	
Impaired assets	

Example:
The exponential of the coefficient gives you the odds ratio, i.e. how large are the odds of having Y=1 relative to before if X increases by one unit.

• The coefficients for the capital ratio is -076
• The estimated odds ratio for capital ratio is $e^{-0.76} = 0.47$
• So for a one unit increase in the capital ratio, the estimated odds of getting in financial distress 2 years later are 53% lower than before, all other risk factors remaining constant.

What is the marginal effect? → No clear answer, marginal effects depend on the level of X

Paper: Fu et al. (2012)

Topic: financial reporting. Method: empirical/archival

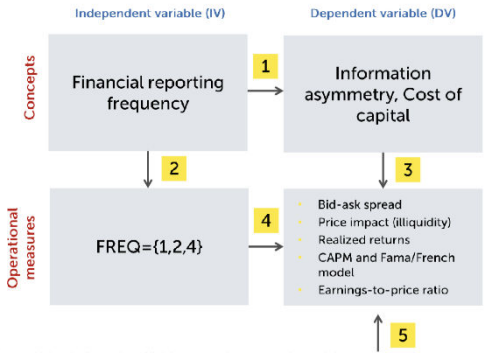
RQ: Does the frequency of financial report affect fair and efficient resource allocation in the economy?
→ Focus on two important aspects: information asymmetry and cost of equity

Regulators require frequent reporting, which imposes costs on companies.
• Benefits for users: better information for forecasting dividends and evaluating firm value.
• Benefits for preparers (firms): better terms of trade when issuing shares.

There is worldwide variation: some countries require only annual reports, others more frequent. Some firms voluntarily report more often.

Importance of the Question
• For regulators: does it make sense to require frequent reporting?
• For managers: do requirements help them?
• For shareholders: are they better off?
• For academics: helps test theories and empirical debates on disclosure, information asymmetry, and cost of capital.

Theoretical Constructs
• Financial reporting frequency: annual, semi-annual, quarterly.
• Information asymmetry: differences in knowledge between firms (managers) and shareholders, leading to problems like adverse selection (one party knows more about the true value than the other, 'lemons problem').
• Cost of equity: reflects compensation investors require for risk. Frequent reporting may reduce this cost, but there could also be an information risk.



Null hypotheses:
H1: Higher reporting frequency does not affect information asymmetry
H2: Higher reporting frequency does not affect cost of equity.

Alternative hypotheses:
H1a: Higher reporting frequency reduces information asymmetry.

H1b: Higher reporting frequency increases information asymmetry.
H2a: Higher reporting frequency reduces cost of equity.
H2b: Higher reporting frequency increases cost of equity.

Independent variable (X): reporting frequency (observable).
Dependent variables (Y): information asymmetry, cost of equity (need proxies).

Table 5
Regression results with information asymmetry measures as dependent variables.

Panel A: Bid-ask spread model

$$IA_{Spread_{i,t}} = \alpha + \beta_1\, Freq_{i,t-1} + \beta_2\, Size_{i,t-1} + \beta_3\, log(Turnover_{i,t-1}) + \beta_4\, log(Volatility_{i,t-1}) + \varepsilon_{i,t}$$

Variable	OLS
Frequency	−0.146***
	(−3.11)
Size	−0.072***
	(−6.85)
log(Turnover)	−0.089***
	(−4.20)
log(Volatility)	1.245***
	(17.81)
Fixed effects	Industry
Adj. R^2	42.97%
F-statistic	81.56***

Results
• Reporting frequency negatively related to bid-ask spreads:
• Coefficient: -0.146 → if you increase reporting frequency by one, then the bid-ask spread is reduced by 0.146 percentage points.
• T-statistic is -3.11. (standard error is -0.1456/-3.11=0.047).
→ Since the t-statistic is -3.11 and that is greater than 2.57 or lower than -2.57, we can reject the null hypotheses at a significance level of 1% and we can accept β_1.

Conclusion: More frequent financial reporting reduces information asymmetry among market participants.

Paper Core, Holthausen & Larcker (1999)

Topic: Corporate governance. Method: Empirical/archival. Paper uses regression techniques.

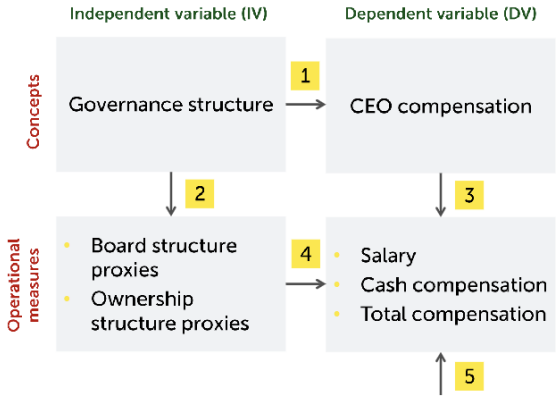
RQ1: Does governance structure influence CEO compensation?
RQ2: Is the relation between governance structure and compensation due to unresolved agency problems?

Governance mechanisms studied: monitoring by board of directors and shareholders.

Theory & Hypotheses

• Agency theory: separation of ownership and control
• CEOs may be (too) powerful due to dispersed ownership, weak monitoring, and control over board information.

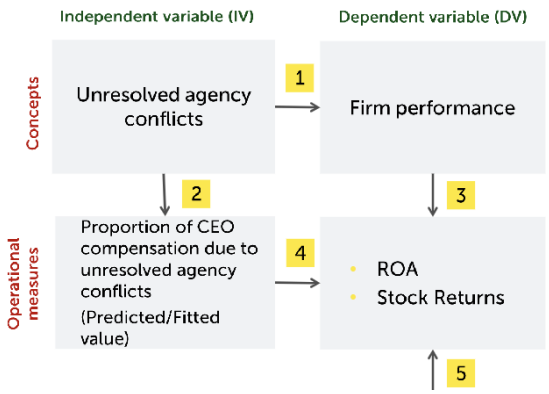
Null hypothesis: Observed board characteristics and ownership structure induce optimal CEO contracting and firm performance. I.e. level of compensation fully explained by economic factors (firm size, complexity, profitability, risk).



Alternative: *less effective* governance structures allows CEOs to capture excess compensation.
• E.g. when the CEO is also chairman of the board, CEO has high ownership, absence of monitoring

Results

• Ownership and board structure variables individually and collectively influence compensation
• But, observed relationship between governance variables and CEO compensation could be due to unresolved agency conflicts (allowing CEOs in weakly governed firms to earn 'excess' compensation) or misspecification of the prediction model for CEO compensation (governance just proxies for other economic characteristics that were not included).
→ Implement 'performance test'



If more compensation is due to agency conflict, unresolved conflicts are costly and lead to lower firm performance.
→ H2: 'Excess' compensation predicts lower firm performance

Predicted values $\hat{Y}$ are values for an output variable Y that have been predicted by a model fitted to the data.

Results:
• Higher excess compensation results in lower stock returns.
• Statistically significant: a one-unit increase in excess compensation reduces stock returns by 12.43 percentage points.
• Economically significant: e.g. if 40% increase in excess compensation, stock returns decrease by $0.4 * 12.43 = 4.97\%$. This is large, because average one-year stock return is 15%

Lecture 5

Recap: Example Nonlinearities

• Looking at a large dataset of earlier audit mandates, a big audit firm wants to investigate the determinants of the necessary size of their auditor teams for different clients.
• You assume that firm size, proxied by sales, is a key determinant of the necessary number of auditors. However, you believe that this relationship is not linear: For small clients, an additional 10 million of sales require more additional auditor resources than for large clients that already have >500 million of sales.

Using the OLS regression framework, how can you incorporate this nonlinear relationship?

→ E.g., adding a squared term: $TeamSize = a + \beta_1 Sales + \beta_2 Sales^2 + \varepsilon$
Or use a lin-log model

Recap: Example Logistic regressions

• For your master thesis, you want to examine the effect of consumer attention on the likelihood of a firm issuing a voluntary ESG report.
• You measure Consumer Attention by counting how often a company is mentioned in newspapers covered by the LexisNexis database.
• ESG is a dummy variable that is 1 when a company issues a voluntary ESG report, and 0 otherwise.
• A friend of yours recommends to apply a logistic regression instead of a standard OLS regression.

Why would a logistic regression be appropriate? And how would you interpret a regression coefficient $\hat{\beta}$ from estimating the logistic regression?

→ When you are estimating the regression with a binary, you are estimating the probability of it happening. Logistic regression would be more appropriate because it limits the outcome range between 0 and 1. And it reflects that when implied probability is already very high, the marginal effect of $Consumer\ Attention$ is lower than when the implied probability is low.
→ $e^{\hat{\beta}}$ indicates the odds ratio (i.e. the odds of issuing an ESG report after increasing consumer attention by one unit relative to the odds before)
→ The marginal effect on the actual probability depends on the level of $Consumer\ Attention$

Correlation, causation, and the counterfactual

Correlation and Causation – Interesting “Insights”
→ Google searches for ‘Taylor Swift’ correlates with Fossil fuel use in British Virgin Islands

Correlation and Causation – The problem
Simply applying the regression methodology is perfectly fine to describe associations between variables. Assuming that circumstances don’t change a lot, associations are also often good enough to make decent predictions.

But we can’t automatically trust correlation to tell us anything about causal effects!
• Remember Core, Holthausen, and Larcker (1999): Agency conflicts drive executive compensation?
• OLS assumption IV. ($E[\varepsilon \mid X] = 0$) will be key here

How can we facilitate credible causal inferences from empirical observations?

Causal Inferences – The Counterfactual

Philosophical problem of induction: How can we infer underlying causal relationships based on empirical observations? I.e., what is the underlying "data generating process"?

Asking about the counterfactual and causal relationships means to ask "what if" questions:
→ What would have happened to the outcome (Y) if there had been no "treatment" (no change in X)?

Unfortunately, the counterfactual is (almost) never directly observable.

It is impossible that X changes and stays constant at the same time for the same observation unit!

A rare exception in accounting research: **Donelson et al. (2013)**: *Discontinuities and Earnings Management: Evidence from Restatements Related to Securities Litigation*
• Consider the "earnings surprise" (ES, also referred to as "analyst forecast error"):
→ ES = Actual earnings reported by company – analyst earnings forecast

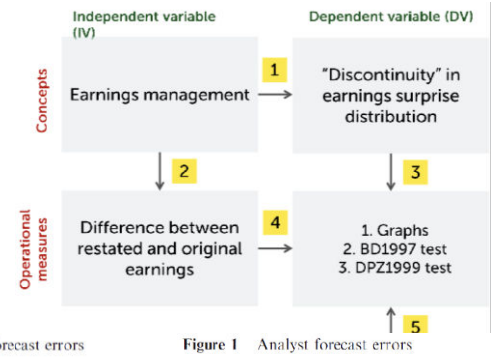
We would expect that for a large sample of analysts and companies over time, ES would have an approximately symmetric proportion of positive and negative values.
• Analysts are sometimes too optimistic, sometimes too pessimistic.
• The expected value of ES is zero, i.e., E(ES) = 0.

However, the frequency distribution based on actual realizations of ES often shows an asymmetric pattern around zero ("discontinuity").

Prior studies have assumed that this phenomenon is explained by earnings management.
• Managers have incentives to use accounting adjustments to make sure that earnings do not miss analyst expectations.
• Missing expectations would result in lower stock price and/or job security.

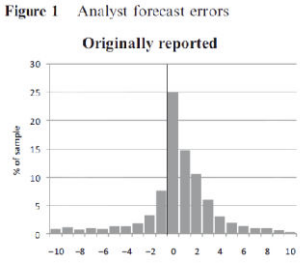
However, it is very difficult to empirically support the assumption of earnings management, because it is itself difficult to observe!

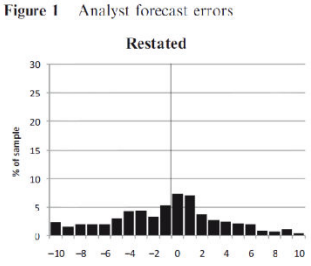
Donelson et al. (2013): Focus on a much smaller sample of companies than prior research.
• Non-random sample: Companies that restate prior financial statements as a result of securities litigation (companies get sued by shareholders).
• Probably not a very representative sample.



But: For this small sample of companies, the study holds constant all other factors except earnings management. Holding all else constant, every company in the sample reports two earnings numbers:
• Earnings that have been manipulated ("originally reported earnings").
• Earnings that have not been manipulated ("restated earnings").

Based on these two numbers, Donelson et al. (2012) are able to calculate the amount of earnings management for every company in their sample:
→ EM = Originally reported earnings – restated earnings





Research Question:
Does earnings management (X) cause the discontinuity in earnings distributions (Y)?

Research Design: Examine the frequency distribution using: Originally reported earnings ("manipulated") and Restated earnings ("true earnings").
• Discontinuity can only change as a result of earnings management.
• All other explanations can be ruled out because nothing else changes!

"Originally reported" graph shows frequency distribution with earnings management. For this sample the discontinuity is clearly visible: many more observations just above zero than just below zero

“Restated” graph shows frequency distribution without earnings management. The discontinuity is smaller for unmanaged (true) earnings. Much smoother distribution around zero, without discontinuity. This provides evidence that the discontinuity in reported earnings is due to earnings management

Conclusion: For this sample of companies, earnings management seems to be causing the discontinuity in the distribution of earnings surprises. Because of the unique research design, we can have relatively high confidence in the causal effect: The only variable that changes is earnings management, and it appears to have a significant effect on the earnings management discontinuity.

The key takeaway is that comparing originally reported vs. restated earnings allows us to observe the counterfactual and shows that discontinuities in reported data are indeed due to earnings management.

Estimating the counterfactual

If we cannot directly observe the counterfactual (i.e., almost always), we need to approximate it: Isolate variation in Y that is (likely) caused by variation in X.

Ideal procedure: randomly assign different levels of the treatment X and then observe the corresponding level of the dependent variable Y.
• It makes it very unlikely that X is correlated to/caused by other factors that cause variation in Y.
• Random assignment is not the same as holding everything constant, but it has the same effect: in expectation, all other determinants of Y should be the same at the different levels of X.
• With random assignment, we can say that whether and how an observation is treated is “exogenous” to its other characteristics. Such a experiment is, e.g., the gold standard for medical investigations.

When independent variable X is exogeneous, we can make inferences about its causal effect on Y!

Obstacles to causal inferences

When using archival (i.e., “real world”) data instead of experiments, there is a high probability that X is not randomly assigned, but instead is “endogenous”, and we have “endogeneity concerns”: in the presence of endogeneity, regression estimations do not measure the true underlying causal effects.

Three sources of endogeneity:

• Omitted variable(s): a relevant variable is not included

Example: Does more analyst following drive higher earnings quality? Maybe, but it could also be driven by firm size: large firms attract more analysts, and large firms also have better accounting.

• Reverse causality / Simultaneity: the "effect" is the cause (Y→X) / a reciprocal relationship (X↔Y)

Example: Does having a Big 4 auditor reduce earnings management? Or are firms with less earnings management more likely to hire Big 4 auditors? Or do both influence each other simultaneously?

• Measurement error in the independent variable (more technicality, not driven by non-random assignment of X).

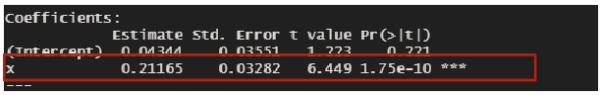
Endogeneity 1: Omitted variables

Where X is not randomly assigned, it is likely correlated with some other characteristics Z, which might also be determinants of Y. In particular, if such characteristics Z themselves (partly) determine both X and Y, they are called confounders.

To isolate the true standalone effect of X, the effect of Z should be “controlled for” by including Z as an independent variable to the regression model. Then β_X gives you the effect of X “holding Z constant”.

Common problem: Sometimes Z is unobservable or hard to measure, or you might not even be aware of its influence. Leaving out Z will then cause “omitted variable bias”.

Results of a regression of Y on X (omitting Z): $Y = \alpha + \beta \times [0.5 \times Z + m] + \varepsilon$



The first regression output table shows that X has a statistically significant coefficient when Z is omitted, even though the true effect is zero.
Once we control for Z, the spurious effect of X on Y disappears.
Second regression table confirms this: when Z is included, the coefficient of X is not significant anymore.

The previous example shows that not controlling for a variable (Z) that is a determinant of both X and Y can bias the coefficient (and the standard error) for X.

Remember: Y = α + βX + ε.
The error term ε captures all factors that affect Y other than X.
Omitting Z from the model means that Z is captured in the error term (because Z affects Y).
Since Z is also causes X, this means that ε is correlated with X.

	$corr(X,Z) > 0$	$corr(X,Z) = 0$	$corr(X,Z) < 0$
$\gamma > 0$	Positive bias of $\hat{\beta}$	$E(\hat{\beta}) = \beta$	Negative bias of $\hat{\beta}$
$\gamma = 0$	$E(\hat{\beta}) = \beta$	$E(\hat{\beta}) = \beta$	$E(\hat{\beta}) = \beta$
$\gamma < 0$	Negative bias of $\hat{\beta}$	$E(\hat{\beta}) = \beta$	Positive bias of $\hat{\beta}$

This is the definition of endogeneity: explanatory variables and error term are correlated: $E(\varepsilon|X) \neq 0$.
…which gives you an intuitive understanding of OLS assumption IV.

When there is correlation between the independent variables and the error term, OLS will not hold.

How bad is it in general when you have endogeneity from omitted variables? It depends…
• On the strength of the relationship between Z and Y.
• And on the strength of the relationship between Z and X.

If X and Z are very weakly correlated, or Z has just minimal influence on Y, you don't need to worry.
Otherwise: you have a problem.

Shows the sign of bias depending on correlation between X and Z (positive, zero, negative) and sign of γ. $Y = \alpha + \beta X + \gamma Z + \varepsilon$

Endogeneity 2: Reverse causality/Simultaneity

X can also be endogenous due to reverse causality/simultaneity:

Example:
Hypothesis: A high-quality auditor causes lower earnings management: $EM = \alpha + \beta \times Big4 + \varepsilon$
But if firms that don't do earnings management voluntarily decide to hire high quality auditors to signal to investors that they have nothing to hide: Hypothesis 2: $Big4 = \eta + \delta \times EM + \gamma$

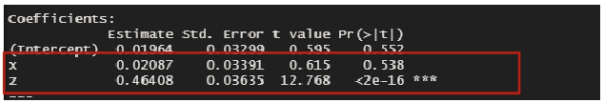
Endogeneity 3: Measurement error

Even a variable that is in principle a good operationalization of a construct is often measured with error: e.g. wrong entries in databases, people who fill out surveys with errors etc.

Then instead of X you measure $\tilde{X} = X + \rho$, with some error ρ.
The true regression model is: $Y = \alpha + \beta\tilde{X} + \tilde{\varepsilon}$

But $\tilde{X}$ and $\tilde{\varepsilon}$ depend on ρ, which makes them correlated (and therefore introduces endogeneity).

Endogeneity from measurement errors in the independent variable always introduces attenuation bias:
$\hat{\beta}$ as an estimate of the true regression coefficient β is biased towards zero.
• As measurement error increases, the slope of the regression line flattens (closer to zero), showing attenuation bias. So, the larger the measurement error, the stronger the downwards bias of $\hat{\beta}$.



Measurement error introduces an endogeneity problem only for the independent variables, for the dependent variable it just disappears in the error term.

Endogeneity – Example

Hypothesis: “Firms with high quality auditors obtain higher valuations in M&A transactions.”

Your measure of firm valuations is the earnings multiple ($MULTIPLE$), i.e., the ratio of the transaction price to earnings. Your measure of audit quality is an indicator variable that captures whether the company is audited by a Big 4 auditor ($BIG4$).

Question: What three sources of endogeneity could be present if you run a simple linear regression?
→ Omitted variables (management/governance quality), but not reverse causality (auditor is fixed) and not measurement error

How to address endogeneity concerns?

- Simply showing more associations that are “in line with” some theory can be informative.
- But overall, producing credible inferences about causal relationships is a challenge of research.

Methods to identify causal relationships are called identification strategies: “identifying” or isolating that part of the variation in the data that describes the causal mechanism in question.

Proper identification is crucial for the publication prospects of a study (see Armstrong et al., 2022).

In academic papers, you will see different identification strategies. There is no fixed canon, but, e.g., some (more or less advanced) strategies are:

- Fixed effects (next lecture)
- Difference-in-differences (next lecture)
- Instrumental variables (lecture 7)
- Matching, simultaneous equation models, regression discontinuity designs, …

Avoid endogeneity: Laboratory experiments

When a lack of random assignment potentially introduces endogeneity, then use random assignment.

Example (Van Rinsum, Maas, Stolker 2018): Randomly assigned participants into two groups:

- Using checklist (treatment group; X=1)
- Not using checklist (control group; X=0)

Assignment due to chance rather than participants’ characteristics.

Random assignment ensures that X is uncorrelated with any Z and X is not caused by Y.

Hence, laboratory experiments allow for making causal inferences, i.e. they have high internal validity.

But what about external validity, i.e., validity of our experimental inferences for the overall population? Might there be sample selection bias?

Observational studies (excl. natural experiments):	Controlled laboratory experiments:
No random assignment	Random assignment
X is often endogenous	X is exogenous
Causal inferences difficult	Causal inferences easy
Lower internal validity	Higher internal validity

Mitigate endogeneity: Natural experiments

Natural experiment: An observational study in which the experimental conditions (treatment versus control) are beyond the control of the researcher, but nonetheless randomly assigned by “nature”.

- Combines the internal validity benefits of an experiment with the external validity benefits of an observational study. For an example, see Irani & Oesch (2013) next week.

Mitigate endogeneity: Control variables

The minimum approach in the absence of (quasi-)experimental settings:
• If you reasonably expect there are other measurable variables Z correlated with X that affect Y, you can enter Z in the regression model as control variables.
• When included as a control variable, you essentially remove the variation in Y caused by Z, so it can't be misattributed to X ("holding Z constant"/"ceteris paribus").
• As such, virtually all archival accounting research studies use real-world data and include references on the relation between X and Y on multivariate rather than univariate regression models.

Standard regression model: $Y = \alpha + \beta \times X + \Sigma\gamma \times Controls + \varepsilon$

In practice, you need to think carefully about the underlying causal structure to decide which variables should be included as controls.

Some scenarios of good controls Z (reduces bias in $\hat{\beta}$):
- Z is a confounder. Social status of a student's parents (Z) influences both whether she gets into a prestigious university (X), and whether she will get a high-profile first job (Y).
- Z blocks a confounder: Parents' social status (U) is unobservable, but we can control for whether the student went to an expensive private school (Z).

Some scenarios of bad controls Z (their inclusion increases bias in $\hat{\beta}$):
- Overcontrol bias: You control away the very effect you want to capture.
  Example: the effect of smoking (X) on early death (Y), controlling for having lung cancer (Z).
- Z is a "collider": Example: the effect of body height (X) on basketball players' speed (Y), but controlling for playing in the NBA (Z). In principle, taller makes you a better basketball player, but you can compensate for being shorter by being fast.

Holding NBA constant, you will observe a negative bias.

Some scenarios of unnecessary controls Z (their inclusion has no influence on bias in $\hat{\beta}$, but might influence its precision (i.e., $SE_{\hat{\beta}}$)):
- Z → Y: Might decrease precision. Z is removed from ε, smaller ε implies smaller standard error of β_X.
- Z → X: Might increase precision. Z increases multicollinearity without removing any bias. Intuitively, you remove useful variation in X.

Lecture 6

Recap:

The counterfactual: A "what-if" question: identifying what would have happened if an observation received a different treatment. Used to assess causal effects.

Endogeneity occurs when explanatory variables are correlated with the regression error term:
→ E(ε | X) ≠ 0. Consequences: Leads to biased regression coefficients and standard errors.

Sources of Endogeneity
1. Omitted variables (e.g., intrinsic abilities, wealth).
2. Reverse causality (Y influences X).
3. Measurement error in X.

Example: Studying effect of university education on startup success.
- Omitted variables: founder's wealth or talent.
- Reverse causality: unlikely (education precedes success).
- Measurement error: possible but minor.

Mitigating endogeneity via Fixed Effects.

Fixed effects: Control variables accounting for unobserved but constant characteristics. Useful for panel data (many entities across time).
- Cross-sectional data: many entities at one time.
- Time-series data: one entity across many times.

Equation: $Y_{ij} = \alpha_0 + \beta X_{ij} + \Sigma\gamma * Controls_{ij} + \rho_i + \varepsilon_{ij}$
Where ρ_i = unobserved effect (constant for each group).

Fixed effects capture:
- Repeated observations of same units,

• Groups (e.g., industries), or
• Time periods (e.g., year effects).

If ρ_i correlates with X → endogeneity problem.

Example: Studying how study time (X) affects test results (Y) across 3 groups (A, B, C).

Regression $Y = \alpha + \beta X + \varepsilon$ shows: "More study time → better grades" (positive slope).

But if groups differ in ability:
• Ability negatively correlated with study time (smart students study less).
• Ability positively correlated with grades.

Then regression shows a negative relationship (downward-sloping line).

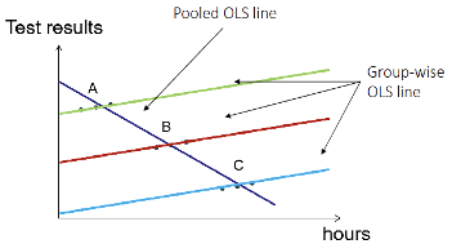
Key idea: Unobserved heterogeneity distorts β.

Introducing Fixed Effects

Introduce group dummies (equivalent to separate intercepts per group) to control for unobserved group characteristics: $Y = \alpha + \Sigma\delta_i D_i + \beta X + \Sigma\gamma Controls + \varepsilon$

This removes between-group variation, using only within-group variation to estimate β.

Now, the regression allows separate baselines (α_i) per group.
"Studying more increases grades, conditional on ability level."



Disadvantages of Fixed Effects

• Cannot estimate effects of variables constant within groups.
• Reduces variation → larger standard errors.
• De-meaning removes between-group variation.
So, use only if sufficient within-group variation exists in X.

If there is little distortion but fixed effects are still applied, β becomes less precise. So, don't use FE unnecessarily; you lose useful variation.

Difference-in-Differences Design

• Goal: Measure the effect of X on Y when a treatment occurs at a specific time.
• Example: a regulation, macroeconomic shock, or stress test.
• Approach: Combine exogenous shocks over time with group-wise fixed effects.

It's the most popular quasi-experimental design today. Armstrong et al. (2022): 65% use DiD.

Example IFRS

Research question: Did the mandatory IFRS adoption in 2005 reduce firms' cost of capital?

Steps:
• Ignore early voluntary IFRS adopters (they're self-selected).
• Focus on companies that mandatorily switched in 2005.
• Compare: Pre-2005 (domestic standards) vs. Post-2005 (IFRS).
• Question: Did cost of capital drop in the post-IFRS period?

A basic univariate test shows:

	Pre 2005	Post 2005	Difference
Avg. Cost of Capital	0.13	0.10	-0.03

→ 3% reduction → seems like IFRS helped.

But can we be sure this is causal? Or just coincidence? → Possible Confounding Factors: Other regulations at the same time, changes in markets or firm characteristics.

We must compare the observed decrease to a counterfactual; What if IFRS had not been introduced?
• Counterfactual = "What if" analysis.

Two situations: Firm *did* report under IFRS in 2005+ | Firm *did not* report under IFRS in 2005+
Goal: compare these two outcomes for the same type of firm.

IFRS Counterfactual Matrix

Conditional outcome	**Pre 2005**	**Post 2005**
Y(IFRS)	Counterfactual	Observed Y
Y(no IFRS)	Observed Y	Counterfactual

Counterfactual I (first column): what if treatment existed before it actually happened.

Counterfactual II (second column): what if no treatment after 2005.

→ DiD approximates Counterfactual II.

Implementing DiD

• Goal: Approximate counterfactual II using a control group. Compare:
  1. Treatment: countries that switched to IFRS.
  2. Control: countries that did not switch.
• Important: Groups need not have identical firms, only stable characteristics. Don't only compare after 2005; base levels may differ.

Examine change in outcome (not level) between groups:
→ IFRS firms (treatment) vs. non-IFRS firms (control).

Core assumption: both follow a parallel trend before treatment.

The parallel trend assumption is critical for DiD validity.

It means: The control group represents what would have happened to treated firms without treatment.

But we can't "prove" it statistically, we assess plausibility.

Arguments for believing parallel trends:
• Treated & control groups should have similar characteristics.
• Better if treatment is quasi-random (e.g., Irani & Oesch 2013) than targeted shocks.
• No reason to expect sudden divergence before treatment.

Visual inspection: Plot outcome over time, do pre-treatment trends look parallel?
• Alternative check: Placebo tests
• Pretend treatment happened earlier → test if any difference arises.
  If none → supports assumption.

Difference-in-Differences Table

Subsample	Pre 2005	Post 2005	Difference (Post–Pre)
IFRS firms	0.13	0.10	-0.03
Non-IFRS firms	0.10	0.11	+0.01
Difference (Treatment–Control)	0.03	-0.01	-0.04

→ Interpretation: IFRS adoption reduced cost of capital by 4%.
→ Larger effect than naive before–after test.

DiD Regression Equation

Regression form: $Y = \beta 0 + \beta 1 Treatment + \beta 2 POST + \beta 3 (Treatment \times POST) + \varepsilon$

• Treatment = 1 for IFRS countries, 0 otherwise.
• POST = 1 for years ≥ 2005.
• β_3 = coefficient of interest → the treatment effect.

Equation applied to cost of capital: $Cost\_of\_Capital = \beta 0 + \beta 1 IFRS + \beta 2 POST + \beta 3 (IFRS \times POST) + \varepsilon$

Should We Add Control Variables? → Yes, DiD can still include control variables because
• Group characteristics might change.
• Parallel trend may only hold conditionally.
• Control variables reduce error variance.

However, DiD already handles unobserved, time-invariant factors; Controls add nuance.

Extensions and Real-World Use
• Simple DiD = single treatment date.
• Staggered treatment: treatment occurs at different times.
• Many studies use two-way fixed effects models (firm & year FE).

But: Recent econometrics shows staggered DiD can be biased → need adjustments.

Landsman et al. (2012)
"The information content of annual earnings announcements and mandatory adoption of IFRS."
Published in Journal of Accounting and Economics (2012).
Authors: Wayne R. Landsman, Edward L. Maydew, Jacob R. Thornock.

Main RQ: Does mandatory IFRS adoption increase the information content of annual earnings announcements?

Topic: Financial Accounting.
Method: Empirical archival analysis using public data.

Follow-up questions:
• Conditions: Moderating factors (when/where is effect stronger or weaker?).
• Mechanisms: Mediating factors (why/how does effect occur?).

Why This Study Matters
• IFRS adoption (2005) was costly for firms and regulators. Need evidence of benefits.
• Earlier studies gave mixed evidence on whether IFRS improved accounting quality.
→ This study aims to provide empirical clarity.

Hypothesis and Operationalization

Key idea: Earnings announcements are important information events (Ball & Brown 1968).

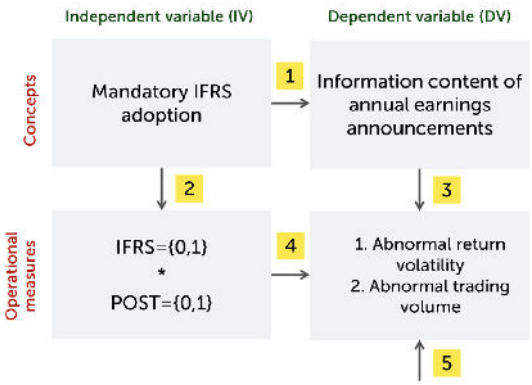
Hypothesis (H1):
Mandatory IFRS adoption increased the information content of annual earnings announcements.

X = Mandatory IFRS adoption (observable).
Y = Information content of announcements (unobservable, needs proxies).

To measure "information content," the authors use stock-market reactions rather than abnormal returns directly: → Trading volume and return volatility serve as proxies for investor response to new information.

How to Measure Y (Information Content)
Two empirical proxies: (Both capture how intensely investors react to new earnings news.)
1. Abnormal Return Volatility (AVAR)
2. Abnormal Trading Volume (AVOL)

Regression Model (Design)



The paper uses a difference-in-differences mode (similarly for AVOL).:

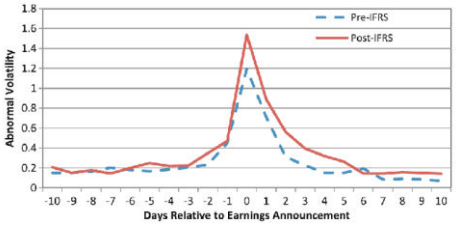
$$AVARit = \beta 0 + \beta 1 IFRSit + \beta 2 POSTit + \beta 3 (IFRS \times POST) it + \Sigma k \beta k CONTROLkit + \varepsilon it$$

IFRS: 1 = IFRS-country firm, 0 = control-country firm.
POST: 1 = post-2005, 0 = pre-2005.
β_3: DiD coefficient = effect of IFRS adoption on information content.

Validity of Parallel Trends

Lists IFRS vs. control countries:
• IFRS countries: Europe (UK, Germany, France, Netherlands …), etc.
• Control countries: U.S., Canada, Japan, etc.
→ Populations are quite different, so parallel-trend validity is questionable.

Raises concern that treatment and control firms differ systematically.



Main Result (H1): Graph: "Daily abnormal return volatility around earnings announcements."
• After IFRS: a much higher volatility spike on announcement day.
→ Greater information release → supports H1.

Numerical Results (H1)

	Non-IFRS (A)	IFRS (B)	Difference (B − A)
Pre-Adoption	–0.02	0.05	0.07 ***
Post-Adoption	0.04	0.30	0.26 ***
Difference (Post – Pre)	0.06 **	0.25 ***	0.19 ***

Interpretation: Abnormal volatility (AVAR) rose 0.25 for IFRS vs 0.06 for non-IFRS.

Difference-in-differences = 0.19 (significant). → IFRS adoption increased information content.

Regression Results
• Dependent variable = AVAR.
• Coefficient on IFRS × POST ≈ +0.18 (significant).
• Even after adding control variables (SIZE, LEVERAGE, LIQUIDITY, etc.), the effect stays positive though smaller.
→ Supports H1.

Conclusions

• Similar results using AVOL (trading volume).
• IFRS countries show a significant rise in the information content of earnings announcements after mandatory IFRS adoption.
• DiD design helps rule out confounding events, strengthening causal interpretation.

Two follow-up analyses:
• Conditions (Moderation): Enforcement quality influences the magnitude of IFRS effects.
• Mechanisms (Mediation): Post-IFRS changes in Reporting lag, Analyst following, Foreign investment.

Irani & Oesch (2013)|

"Monitoring and corporate disclosure: Evidence from a natural experiment." Irani and Oesch (2013), Journal of Financial Economics.

Research Question and Design

RQ: Does external monitoring by financial analysts lead to better financial reporting quality?

- Topic: Financial Accounting / Corporate Governance.
- Method: Empirical archival data.
- Follow-up: Does this effect depend on other governance mechanisms?
  "Weak shareholder rights" → substitution between analyst monitoring and internal governance.

Why This Study Matters

- Managerial misreporting ("cooking the books") is costly and distorts markets.
- Analysts help detect and discipline management.
- However, relationship is endogenous; firms with good reporting attract more analysts.
  → We can't just compare firms with many vs. few analysts.
  Need a causal setup to break endogeneity.

Hypothesis and Operationalization

H1: More analyst following → higher financial reporting quality (FRQ).

X: Analyst monitoring (from I/B/E/S databases).
Y: Financial reporting quality → proxied by abnormal (discretionary) accruals.

- Lower discretionary accruals = better FRQ.
- Alternative proxies: total accruals, current accruals, readability ("fog index"), and word count.

Identification Strategy (Idea)

- For causal inference, you want the independent variable X to be exogenous. If it's not → endogeneity
- Solution: Brokerage-house mergers → natural reductions in analyst coverage.
- Analysts work for brokers; when two brokers merge, duplicate coverage is cut.

Example: If Company A was covered by Broker K and Broker L, and K & L merge → one analyst loses assignment.→ Reduction in analyst coverage is "plausibly exogenous."

Exogeneity of Shock

- Quote: "Coverage termination is independent of unobservable firm characteristics."
- Thus, reduction is exogenous to company A's traits.

Natural Experiment Definition

Recalls definition (from last week): A natural experiment = observational study where conditions (treatment vs. control) are beyond researcher's control but assigned by nature.
→ Brokerage mergers fit this perfectly:

- Authors don't control the mergers.
- Assignment to treatment is "as if random."

This setting is even better than a regulatory shock like IFRS, because treated and control firms come from the same overall population.

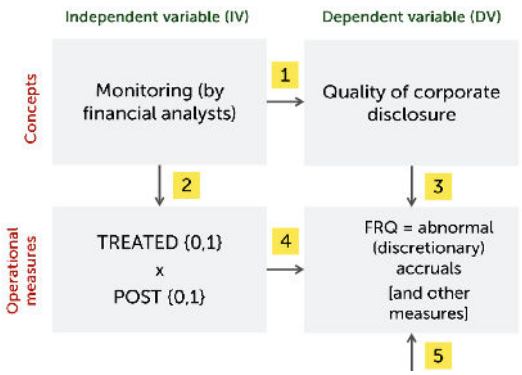
Empirical Model (Difference-in-Differences)

$$FRQ = \alpha + \beta 1 POST + \beta 2 TREATED + \beta 3 (POST \times TREATED) + [\dots] + \varepsilon$$

Where:

- FRQ = financial reporting quality measure (dependent variable).
- POST = 1 if brokerage merger.
- TREATED = 1 if firms affected by coverage drop (treatment group).
- POST × TREATED = main variable of interest (the causal effect).
- "[…]" = firm controls and fixed effects.

Interpretation: β_3 (coefficient of interest) tells how FRQ changed after coverage dropped.



Key additions:
• There were not just one, but 13 merger events studied.
• The data structure is complex and not entirely transparent.

For each event:
• Pre-period: one year before merger
• Post-period: first full fiscal year after merger

• Treatment firms: only show two observations before and after merger.
• Control firms: not affected by any merger.

Controls include: Baseline variation among treatment and control firms. Industry and firm fixed effects.

Results

Column 1 (Y = COVERAGE): Analyst coverage drops by 1 analyst on average for treatment firms post-merger (relative to control group).

Columns 2–5 (Y = FRQ):
POST × TREATED = +0.030 to +0.027 (significant at 5%)
FRQ increases by ~0.03 for treatment firms post-merger.

	COVERAGE	*FRQ*	*FRQ*	*FRQ*	*FRQ*
POST	0.426***	−0.002	0.001	0.000	0.002
	(2.818)	(−0.158)	(0.100)	(0.000)	(0.118)
TREATED	16.950***	−0.048***	−0.044***	−0.020***	−0.005
	(21.617)	(−3.811)	(−3.839)	(−3.165)	(−0.765)
POST × TREATED	−1.120***	0.030***	0.028**	0.027***	0.025**
	(−3.717)	(2.420)	(2.346)	(2.182)	(2.126)
Merger fixed effects	No	No	Yes	Yes	Yes
Industry fixed effects	No	No	No	Yes	Yes
Firm fixed effects	No	No	No	No	Yes
Number of observations	93,005	93,005	93,005	93,005	93,005
R-squared	0.111	0.001	0.034	0.190	0.419

So, financial reporting quality decreases (since higher FRQ = more discretionary accruals).

Discussion points:
• When applying firm-fixed effects, how can industry fixed effects also be included? Maybe some firms switch industries.
• Estimation issues: Can you estimate TREATED alongside firm fixed effects? Do treatment firms from one merger act as controls for others? Some mergers occur in the same year → overlap between treatment/control.

Conclusion: Analysts do monitor managerial behavior. A drop in coverage → lower disclosure quality.

Additional insights: Moderating effect: Stronger in firms with poor corporate governance.

Fixed effects summarized:

Fixed effects (FE) are a statistical technique that control for unobserved factors constant within groups (e.g., firm, person, country) but differing across groups. They help isolate within-group variation and reduce bias from unobserved heterogeneity. $Y_{it} = \alpha + \beta X_{it} + \rho_i + \varepsilon_{it}$

FE models remove unobservable group-specific constants by demeaning data within each group. This isolates how changes within an entity affect the outcome, e.g., how a firm's strategy change affects its own performance.

Advantages: Reduces Endogeneity, Focus on Within-Group Effects, Avoids Omitted Variable Bias (if bias is constant), Widely Used in Panel Data

Disadvantages: Cannot Estimate Effects of Variables That Don't Vary Within a Group, Reduces Variation and Statistical Power, Over-Control Can Hurt, Risk of Multicollinearity

Concept	Meaning / Implication
Goal	Control for unobservable, constant factors within groups
Focus	Within-group (not between-group) variation
Helps with	Endogeneity from omitted variables
Hurts when	Overused → loss of variation, higher SE
Rule of thumb	Use FE only when there is meaningful variation within groups and when unobserved heterogeneity might bias results