Econometrics & Financial Markets

Other tools and methods :
ANOVA, ANCOVA
Logit model
Panel data
Diff-in-diff
Event study

Toulouse Business School
MSc BIF

Anna CALAMIA
a.calamia@tbs-education.fr

ANOVA, ANCOVA

Categorical or qualitative variables

- ANOVA (analysis of variance):
  - Compares means among groups, based on a numerical response variable (dependent variable) and qualitative explanatory variable (*factors*).
  - Seeks to identify sources of variation in the response variable: Variation in DV about its mean is explained by one or more categorical independent variables or is unexplained (random error).
  - Assumptions on errors as in linear regression
  - One-way and Two-way ANOVA

- ANCOVA (analysis of covariance):
  - Similar to ANOVA but uses both qualitative (*factor*) and quantitative (*covariate*) explanatory variables. The variance of the dependent variable is decomposed in variance explained by the covariates, by the factors and the residual variance.

TUTORIAL XLSTAT

7. ANCOVA

- Regression with quantitative and qualitative variables

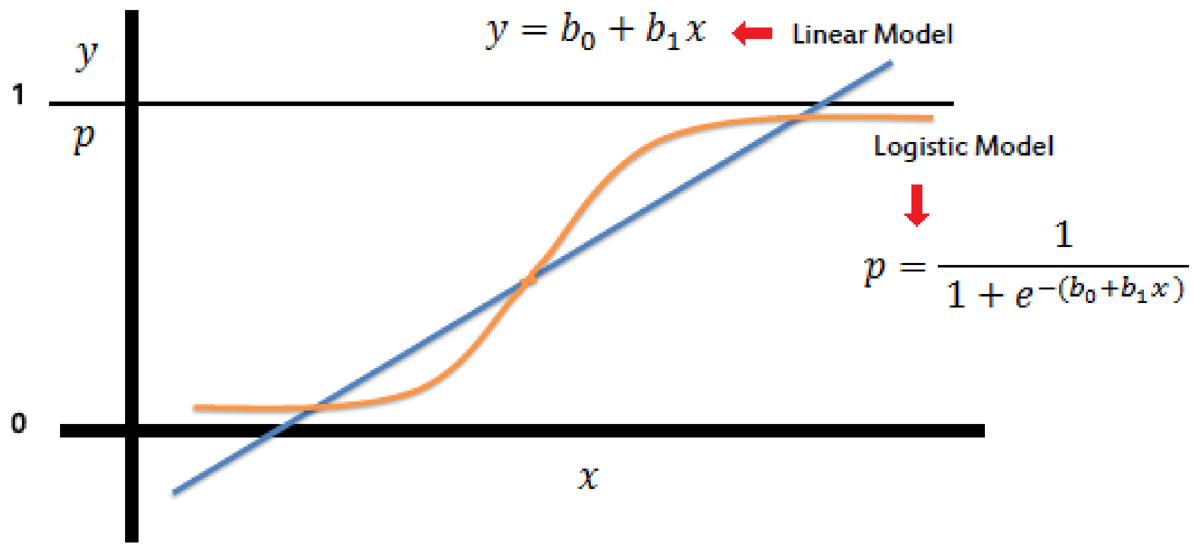
Logit model

Regressions with qualitative dependent variables

- **Qualitative dependent variables** (or categorical dependent variables) are dummy variables used as dependent variables.
- **Probit and logit** regression models are used to model the effect of a series of variables on a **binary response variable** (with two possible values, such as pass/fail)
- They are based on the estimation of the probability of a discrete outcome given the values of the independent variables used to explain that outcome: probability that Y = 1 (a condition is fulfilled) given the values of the independent variables.
- **Probit model** is based on the normal distribution
- **Logit model** is based on the logistic distribution

Logit model

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \varepsilon$$



TUTORIAL XLSTAT
8. LOGIT

- Regression with binary response variables

Panel data analysis

Panel data

Panel Data: observations about different cross sections over time (2 dimensions)

- Y_{it} is observed for all individuals i=1,…,N across all time periods t=1,…,T
- E.g., relation between returns and earnings for several stocks over time
- Panel data allows to control for unobservable variables and heterogeneity

Quarter	Stock	EBIT	Return
2017-Q1	AIRBUS	533	-0.00433
2017-Q2	AIRBUS	529	-0.00305
2017-Q3	AIRBUS	414	0.005125
2017-Q4	AIRBUS	878	-0.01073
2018-Q1	AIRBUS	168	-0.00192
2018-Q2	AIRBUS	871	0.02181
2018-Q3	AIRBUS	1524	-0.00661
2018-Q4	AIRBUS	2155	0.002388
2017-Q1	CARREFOUR	1385	0.005002
2017-Q2	CARREFOUR	546	0.00317
2017-Q3	CARREFOUR	546	0.008257
2017-Q4	CARREFOUR	427	0.001944
2018-Q1	CARREFOUR	427	0.003275
2018-Q2	CARREFOUR	-169	-0.00431
2018-Q3	CARREFOUR	-169	-0.00151
2018-Q4	CARREFOUR	931	0.005734

- **Pooled OLS**
  - homogeneous panel data
  - model parameters are common across individuals: $y_{it}=\alpha+\beta X_{it}+\epsilon_{it}$

Heterogeneous panel: FE and RE

- **Fixed Effects**

  Includes unobservable individual-specific and/or time-specific effects, possibly correlated with the observed explanatory variable:

  $Y_{it} = \beta X_{it} + \alpha_i + \epsilon_{it}$

  $Y_{it} = \beta X_{it} + \alpha_i + \tau_t + \epsilon_{it}$

  where αi is the unknown intercept for each entity (i=1….N), composed of a constant intercept and an individual-specific term; τt captures any unobservable time-specific effects.

  → Within estimator or least squares dummy variable(LSVD)

- **Random Effects**

  Includes unobservable time-specific and/or individual-specific effects which act like individual-specific stochastic error terms, uncorrelated with the regressors.

  → GLS with appropriate error structure (accounting for individual-specific error)

- **Hausman test** to chose between fixed or random effects.

  H0: the preferred model is random effects; H1: fixed effects

  Tests whether the unique errors are correlated with the regressors (H0: they are not)

Panel Data – example

firm	year	inv	value	capital
1	1935	317.6	3078.5	2.8
1	1936	391.8	4661.7	52.6
1	1937	410.6	5387.1	156.9
1	1938	257.7	2792.2	209.2
1	1939	330.8	4313.2	203.4
1	1940	461.2	4643.9	207.2
1	1941	512	4551.2	255.2
1	1942	448	3244.1	303.7
1	1943	499.6	4053.7	264.1
1	1944	547.5	4379.3	201.6
1	1945	561.2	4840.9	265
1	1946	688.1	4900.9	402.2
1	1947	568.9	3526.5	761.5
1	1948	529.2	3254.7	922.4
1	1949	555.1	3700.2	1020.1
1	1950	642.9	3755.6	1099
1	1951	755.9	4833	1207.7
1	1952	891.2	4924.9	1430.5
1	1953	1304.4	6241.7	1777.3
1	1954	1486.7	5593.6	2226.3
2	1935	209.9	1362.4	53.8
2	1936	355.3	1807.1	50.5
2	1937	469.9	2676.3	118.1
2	1938	262.3	1801.9	260.2
2	1939	230.4	1957.3	312.7
2	1940	361.6	2202.9	254.2
2	1941	472.8	2380.5	261.4
2	1942	445.6	2168.6	298.7
2	1943	361.6	1985.1	301.8
2	1944	288.2	1813.9	279.1
2	1945	258.7	1850.2	213.8
2	1946	420.3	2067.7	132.6
2	1947	420.5	1796.7	264.8
2	1948	494.5	1625.8	306.9
2	1949	405.1	1667	351.1
2	1950	418.8	1677.4	357.8
2	1951	588.2	2289.5	342.1
2	1952	645.5	2159.4	444.2
2	1953	641	2031.3	623.6
2	1954	459.3	2115.5	669.7
....	....	....	....	....

Summary statistics:

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
inv	200	0.930	1486.700	145.958	216.875
value	200	58.120	6241.700	1081.681	1314.470
capital	200	0.800	2226.300	276.017	301.104

Results for variable inv:

Goodness of fit statistics:

rsq	0.769
adjrsq	0.767

R-sq:
within = 0.7668
between = 0.8196
overall = 0.8061

Joint test of significance (F or Chi-square test):

statistic.Chisq	parameter.df	p.value.Chisq
657.295	2	1.8634E-143

Coefficients:

	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	-57.865	29.393	-1.969	0.049
value	0.110	0.011	10.429	<0.0001
capital	0.308	0.017	17.948	<0.0001

hausman fixed random
Test: Ho: difference in coefficients not systematic
Prob>chi2 = 0.3119

Difference-in-differences

Difference-in-differences

Used to estimate the effects of a sudden change in economic environment, policy, or general treatment on a population

- **Treatment group**: subject to the change i.e. to the **treatment (**sudden exogenous source of variation)
- **Control group**: similar in characteristic to the treatment group but not subject to the change
- Quantifiable and measurable **outcome**
- Measure of treatment effects based on **between-group cross-sectional differences** and **within-group time-series differences**
- **Parallel trend assumption**: in the absence of treatment, the difference between the groups is constant over time

Example: Card and Krueger(AER,1994)

- Does an increase in minimum wage have a negative impact on employment?

- Study the evolution of the number of employees in fast-food restaurants in New Jersey (NJ) following the increase in minimum wage from $4.25 (Feb. 1992) to $5.05 (Nov. 1992)
- Comparison with the evolution of employment in Pennsylvania(PA), a neighbouring state

Diff-in-diff: basic principle

- The average (expected) number y of employees, in state s at time t:

$$E(y|s,t) = \gamma_s + \lambda_t$$

- Where γs is a constant specific to state s and λt is a constant specific to time period t
- A change (treatment) on minimum wage occurs in state s at period t and creates a shock on employment equal to β
- For a given restaurant i operating in state s at date t, the number of employees will be equal to

$$y_{ist} = \gamma_s + \lambda_t + \beta D_{st} + \epsilon_{ist}$$

- Where ϵist is the error term and Dst is a dummy variable that takes on the value 1 for observations from the treated group (NJ) after the treatment (post), 0 otherwise

Diff-in-diff: basic principle

- The question is: How can we measure β ?
- Compute first the difference between the average number of employees in PA, after and before the treatment
  - ➔ This difference allows to remove differences across states

- Compute next the difference between the average number of employees in NJ, after and before the treatment

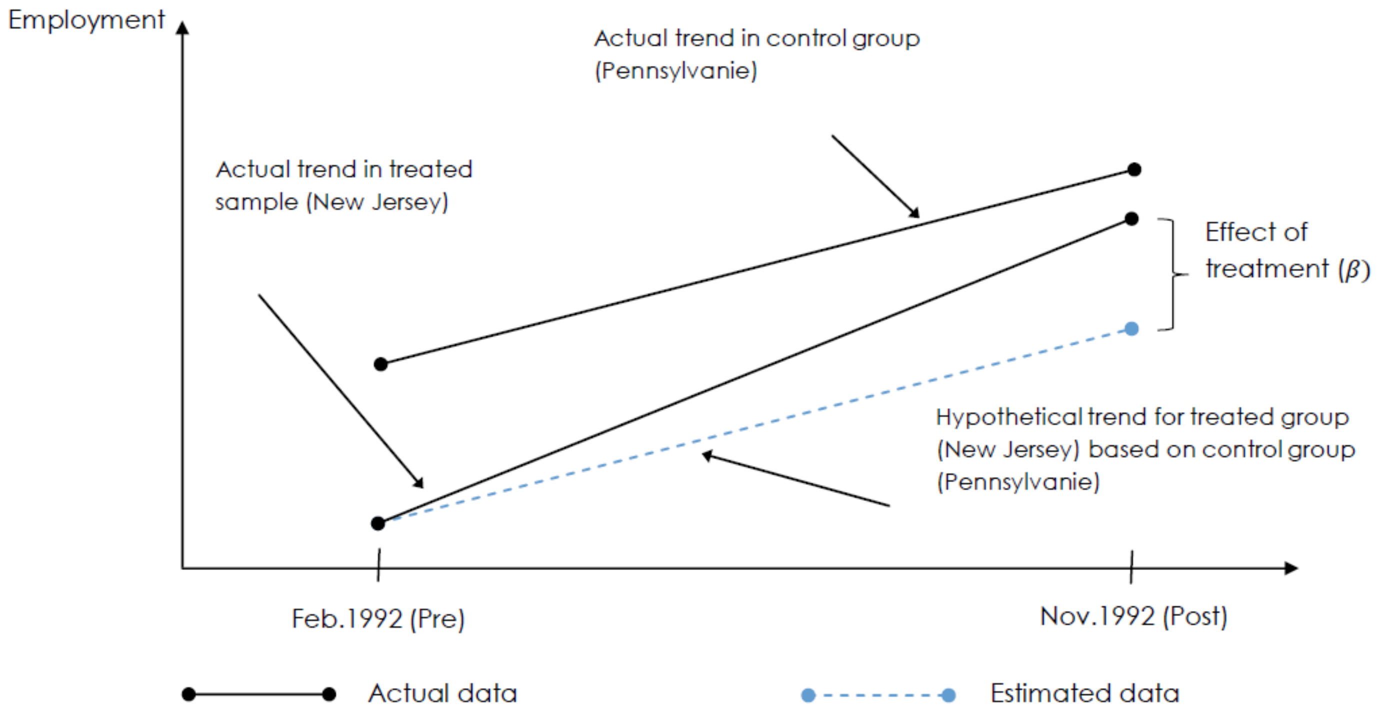
- Finally, compute second difference minus first difference:

$$E(y_i|s = NJ, t = Post) - E(y_i|s = NJ, t = Pre) \quad - \quad [E(y_i|s = PA, t = Post) - E(y_i|s = PA, t = Pre)]$$

➔ This is equal to : β

We have eliminated the common trend between the groups, λt, and the permanent differences between the groups, leaving a very simple estimate of the treatment effect, β.

Diff-in-diff: basic principle



Diff-in-diff: regression specification

- Simple DID does not compute the statistical significance of the shock β
- Regression specification allows to overcome this problem:

$$Y_{ist} = \alpha + \gamma D_s + \lambda D_t + \beta(D_{st}) + \varepsilon_{ist}$$

where: Yist is the number of employees in restaurant i in state s at period t; Ds is a dummy variable that takes on the value 1 for restaurants in NJ (i.e. treated group) and 0 otherwise; Dt is a dummy variable that takes on the value 1 for observations made after the wage increase (treatment); Dst is a dummy (interaction) variable that takes on the value 1 for observations in NJ after the treatment; εist is the error term

What is the interpretation of the regression coefficients ?

α (intercept) is the average number of employees in restaurants operating in PA (during the Pre period)
γ is the difference between the average number of employees in NJ and PA
λ is the difference between the average number of employees working in restaurants in the $Post$ and the Pre periods
β is the DiD estimator, Average differential change in y from the first to the second time period of the treatment group relative to the control group

Event studies

Event studies: what for?

Event studies aim at quantifying the effects of an (unexpected) economic event on the value of firms

- Financial economics: corporate events, market efficiency
- Macroeconomic policy: fed rates, trade deficits
- Accounting: earning announcements
- Law and economics: changes in legal environment and regulation
- Marketing: brand strategy announcements
- ...

How asset prices react to a given event:

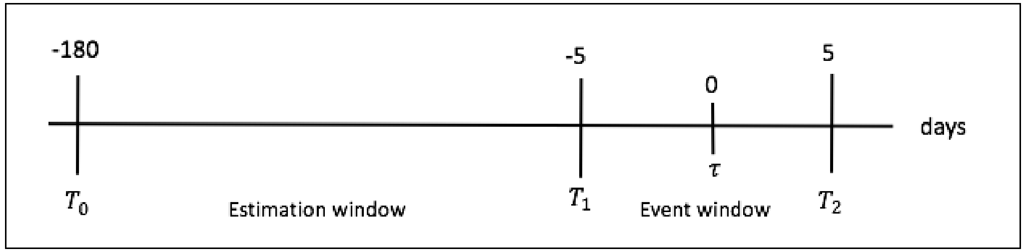
- events are reflected in asset prices (assuming markets are informationally efficient)
- prices are easily observed
- well-performing models are available to isolate the impact of a given event on asset prices

Test of market efficiency: Is the relevant information impounded into prices immediately or with delay?

Event studies design

Step 1: Definition of the event and event window

- Unexpected event
- Exact date (date 0)
- Short vs long horizon
- Define an event window (period over which prices are examined) that is larger than the exact period of interest
  - Inclusion of the days prior to the event (-1, -2, ...) aims at accounting for possible anticipation of the event as well as information leakage
  - Inclusion of the days after the event (+1, +2, ...) aims at capturing posterior abnormal movements that occur after market close



Step 2: Selection criteria (Which firms to be included in the study?)

- Restrictions imposed by data availability and reliability
- Restrictions imposed by representativeness issues
- Some summary statistics (market capitalisation, average return, industry representation, distribution of events through time...) might prove useful to identify potential biases in the initial sample as well as outliers

Step 3: Normal and abnormal returns

- Problem: how to isolate price movements induced by the event of interest from contemporaneous movements, unrelated to the event?

➔ We need a measure of abnormal returns:

Abnormal returns are computed as the difference between observed returns and normal returns (predicted returns):

$$AR_{it} = R^*_{it} - E(R_{it}|X_t)$$

Normal returns correspond to the expected returns if the event had not taken place

➔ market model or other asset pricing models (CAPM, FF 3 factors, etc.) to estimate normal returns: $R_{i,t} = \alpha_i + \beta_i R_{m,t} + \xi_{i,t}$

➔ Cumulative Abnormal Returns (CAR), computed as the cumulative sum of abnormal return over the event window (for a given security)

➔ AR averaged in the cross section of sample stocks to compute AARs on each event date: time series of average abnormal returns (AAR)

➔ AAR across securities can also be aggregated over time to compute Cumulated Average Abnormal Returns over the event window (CAAR)

Step 4: Estimation procedure

- Estimate the parameters of the model that is used to generate normal returns over the event window
  - ➔ The estimation is performed on the **estimation window**
- Check that the estimation window is not contaminated by events that are likely to impact the parameters of the model that generates normal returns
- The **event window** (or part of it) should not be included in the estimation window (when feasible). May also introduce a buffer zone between the estimation window and the event window.
- Common choices for the length of the estimation window are 120 days or 250 days
- The estimation of the parameters can be made through **OLS**

Step 5: Test procedure

Once abnormal returns are computed and aggregated, the objective is to test their significance:

- Test of the null hypothesis: Event has no impact on returns, i.e., no abnormal returns

  - Comparison of the distribution of actual returns with the distribution of predicted returns
  - Typically, the specific null hypothesis to be tested is whether the mean abnormal return in the event window is equal to 0
  - Occasionally, other parameters of the cross-sectional variation in abnormal returns can be used, such as the median or variance
  - Parametric tests, such as t-test (based on normality assumption) and non-parametric tests

Event study results
Interpretation and conclusions

Question the reliability of results:

- Interpretation of results
- Robustness tests using various sub-samples
- Incidence of outliers?
- Sensitivity to the choice of the estimation window?
- Sensitivity to the normal-return generating model?
- Other issues (clustering, event induced variance, partially anticipated events, event-date uncertainty, short vs long horizon…)