Goodness of Fit: R^2= SSE/SST. For SLR, $R^2=(r)^2$| **DF for sample variance and MLR (equations):** We lose 1DF for every x incl. β_0. For sample variance of simple regression, square root of SSR/(n-2). For sample variance in MLR, variance of error/SST(1-R^2), SST = $\sum(x_j\text{-avg}(x)_j)^2$

Figure 1- Transformations (Level-log etc.)

Model	Dependent Variable	Independent Variable	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
Log-level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Interaction term: Using interaction terms, we can quantify how multiple features work together to influence the price. This helps us identify which combinations of features drive the most value – valuable information for crafting pricing strategies and promotions. $Y= \beta_0 + \beta_1 x_i + \beta_2 x_2 + \beta_3(x1^* x_2)+\epsilon$, If **β3 > 0**, it indicates a **synergistic effect**: x_i and x_2 reinforce each other's impact on Y. If **β3 < 0**, it indicates a **dampening effect**: : x_i and x_2 weaken each other's impact on Y. If **β3 = 0**, it means there is no interaction between x_i and x_2, and their effects are additive.

Changing units: Changing the units of measurement of one or more variables does not change the regression itself. It changes only the magnitude of the coefficients. If model has log in it, no changes as it is %. SE changes tho!

Calculations: Predictions: Plug in β's . Quadratic terms (finding turning point): $Y\hat{} = \beta_0 + \beta_1 X + \beta_2 X^2$, turning point: $X^* = -(\beta_1/2\beta_2)$. **Changing units in log-level model**: $\ln(Y) = \beta_0 + \beta_1 X + \epsilon$, a 1-unit change in X leads to a $\beta_1 \times 100\%$ change in Y. **Residual calculation**: $e_i = Y_i - Y\hat{}_i$

[**Bias**: MLR1, MLR2, MLR3, MLR4. | **Inference**: MLR5, MLR6 (in addition to MLR1–MLR4)]

Assumption	Cause	Diagnose	Handle	R Code
MLR1: Linearity. The model is linear in parameters. *(Linearity = residuals have mean zero for every level of the fitted values and predictors)*	(1) Incorrect model specification, (2) missing interaction terms or (3) non-linear relationships	(1) Plot residuals vs. fitted values- should show no pattern. (2) Use scatterplots of independent vs. dependent variables. (3) RESET test (F-test compares long model to smaller models & adds powers of the fitted values (e.g., squared, cubed) to check for functional form misspecification. A significant result indicates nonlinearity. "H_0: The model is well-specified."	(1) Check other assumptions (2) Return to data processing (3) Log transformations / quadratic terms (4) Interaction terms (5) Alternative models	(1) plot(model_name, which=1) (2) resettest(model_name, power = 2:3, type = "fitted")
MLR2: Random Sample. The data is randomly sampled from the population, ensuring that the sample represents the population.	(1) Non-representative or biased sample (2) Sampling errors or convenience sampling, (3) Missing values, outliers, non-random samples Outliers: Types: (1) Extreme values (far outside typical range, often in Y variable), means high residuals, (2) Leverage points (unusual X values). (3) Influential points (observations significantly affecting model estimates (predicted values & parameter estimates) Extreme + Leverage	(1) Inspect data collection methods, (2) Check descriptive statistics for sample characteristics. (3) N/As at random (no bias but, lose data, ↑ uncertainty, means ↑ SE), (4) N/As NOT at random means bias (endogenous). *"In small samples, outliers can have a large effect on model estimates, distorting predictions and coefficient values. Impact of outliers diminishes as N increases."* (5) Studentized residuals (standardizing residuals) to identify outliers. Or external res stud (w/o high leverage observation	(1) Over/under representation: Weighting or post stratification (2) N/As: (a) Listwise delete (stats power down), (b) mean/median imputation (↓ variance->bias), (c) regression imputation (when NAs not random, overstates confidence in predictions), (d) flagging NA's (combines imputation + binary variable | when NA's carry value, may introduce bias), (e) Missing category (for categorical variables -> unknown/undisclosed). (3) Outliers: (a) cap extreme values (winsorizing), (3) add, drop, transform variables, (4) fix data entry issues, (5) remove observations outside intended population.	(1) dataset %>% mutate(res_stud = rstudent(m1), res_stud_large = ifelse(between(res_stud,-3, 3), "Normal", "Extreme")) (2) Leverage: Cooks distance (residuals vs leverage) : plot(m1, 5)
MLR3: No perfect collinearity. The independent variables are not perfectly correlated, so the model can estimate coefficients uniquely.	(1) If independent variables are perfectly correlated, (2) if N is too small in relation to parameters, we need at least k + 1 observations to estimate k + 1, (3) dummy variable trap (e.g., including all categories of a categorical variable). *"High multicollinearity reduces the reliability of individual predictors, making it hard to assess their true relationship with the outcome variable."*	(1) Calculate Variance Inflation Factor (VIF). High VIF (e.g., >10) signals potential multicollinearity. Measures how much variance of a regression coefficient is inflated due to multicollinearity with other predictors. (2) Look at correlation matrices.	(1) Combine highly correlated variables / drop one, (2) increase sample size to help disentangle variable effects, (3) A high degree of linear relationship between x1 and x2 can lead to large variances for the OLS estimators. *"What ultimately matters is how big ˆβj is in relation to its standard error."*	(1) numeric_vars <- data %>% select_if(is.numeric) correlation_matrix <- round(cor(numeric_vars, use = "complete.obs"), 2) print(correlation_matrix) (2) vif(model_name)
MLR4: Zero conditional mean. The error term has an expected value of zero conditional on the independent variables. This is critical for unbiasedness, as it implies no omitted variable bias or endogeneity. *Biggest threat to causal modelling: But time periods do not have to be adjacent in panel data*	(1) Omitted variable bias, (2) Measurement errors (a) in Y, if nonrandom, different slope, (b) in X, worse(different slope either way)!, (3) Reverse causality (simultaneity) -> if x and y are simultaneously correlated are correlated with error term. *"Correlation between 1 'x' and 'e' results in ALL estimators being biased. Overspecifying: not unbiased, but variance of estimators increases."*	(1) Perform a residual analysis, (2) use theoretical reasoning to identify potential omitted variables, (3) CEV assumption: e is unrelated to true x, Cov(x*,ME)=0. CEV deals with random ME in X.	(1) Only think and explain if. + or – bias), (2) determine if CEV assumption holds, then attenuation bias: βx_j is pushed towards 0, we underestimate magnitude. $Cov(x,u-\beta_1 e) = -\beta_1 Cov(x,e) = -\beta_1 \sigma^2_e$: Covariance between x and composite error.	(1) plot(model, which = 1) (2) numeric_vars <- data_set[sapply(data_set, is.numeric)] # Select only numeric columns cor_matrix <- cor(numeric_vars, use = "complete.obs") # Compute correlation matrix print(cor_matrix)

MLR5: Homoskedasticity. The variance of the error term is constant across all levels of the independent variables. This ensures valid standard errors for hypothesis testing and confidence intervals. If it breaks: *OLS ≠ BLUE(best linear unbiased estimator) , large N does not help.*	(1) Presence of heteroskedasticity (non-constant error variance) -does not affect the unbiasedness or consistency of OLS estimators, nor the interpretation of R^2. Hypothesis testing does not hold. Variance of estimators becomes biased. SE and CI are not reliable, invalid t and f stats, (2) Outliers or data skewness.	(1) Breusch-Pagan (BP) or White test, (2) plot residuals vs. fitted values (look for patterns) and (3) scale-location plot (horizontal red line indicates homoscedasticity) *"Fan or funnel shape suggests violation of MLR5, curved patterns or distant points indicate issues with model fit or variance consistency. "*	(1) Robust standard errors, (2) transformations to reduce variance (3) remove Outlier / cap a predictor	(1) BP test or white test (for linear and nonlinear): bptest(model, ~fitted(model) + I(fitted(model)^2)) (2) Robust SEs: cov1 <- vcovHC(model, type = "HC3") robust_se <- sqrt(diag(cov1))
MLR6: Error term is normally distributed. The error term follows a normal distribution. Important for small samples to ensure that hypothesis testing, and confidence intervals are valid (by the Central Limit Theorem, normality is less critical for large samples).	(1) Small sample size. (2) Extreme outliers. (3) Skewed or non-normal error distribution.	(1) Plot a Q-Q plot of residuals: compares distribution of the residuals from regression model to normal distribution. Points are roughly along 45-degree line. Curves suggest skewness. S shaped mean heavier or lighter tails than Dnorm, (3) Perform a Shapiro-Wilk or Kolmogorov-Smirnov test. (if residuals deviate from DNorm	(1) transformation, (2) larger sample	shapiro.test(rstandard(model)) H_0 = Residuals follow a DNorm. **test is appropriate for n<= 5000**

Pooled Cross Section two or more random samples from a population of interest at different points in time. Because samples are collected independently of each other and randomly they need not be of equal size, and they will typically contain different statistical units. **Pooled Time Series** multiple statistical unit of analysis N= {1,...,n}observed in several successive periods of time T= {1,..,t}(e.g. the price of several stocks over time). This differs from panel data in that there are more time periods than statistical units (t>n).

Aspect	Pooled Time Series (Panel)	Pooled Cross Section
Data Structure	Same entities tracked over time	Different entities sampled at each time point.
Time Dependency	Time dependency within entities	No dependency between time points
Example Analysis	Panel data models, fixed/random effects	Difference in means tests, trends in averages
Use Case	Studying individual level trends	Studying population level changes

DIFF in DIFF: A treatment group affected by an intervention, a control group unaffected by the intervention, pre-and post-intervention data for both groups.

Strengths: Provides **intuitive interpretation** of causal effects. Allows use of **observational data** if assumptions (e.g., parallel trends) hold. Works with both **individual- and group-level data**. Focuses on **changes over time**, not absolute levels, accounting for pre-existing differences. Adjusts for **time-varying factors** unrelated to the intervention. Limitations: Requires **baseline data** and a valid **control group**. Treatment allocation depends on the **baseline outcome**. Groups exhibit **different outcome trends** over time. Group **composition changes** pre/post intervention. Note: Units do not self-select into treated/not treated groups. Parallel trends assumptions: In the absence of treatment, the difference in outcomes between the treatment and control groups would remain constant over time. **Unbalanced data** is no problem, if it is not systematic (MLR2). **Idiosyncratic errors (v_{it})** are individual-specific errors in a regression model that vary randomly across observations and over time, capturing factors not explained by the model. **Heterogeneity bias**-> correlation between unobserved effects and explanatory variables.

FD: One period of data drops automatically. For T=2, FD reduces to simple cross section. When panel is balanced and T=2, FD = FE. For T>2, FE for long term effects, FD for short term impact on changes. $y_{it} = \beta_0 + \beta_1 x_{it} + a_i + v_{it}$ -> **unobserved effects (a_t) model**. FD: $(y_{i(t-1)} - y_{it}) = \Delta y_{it} = \beta_1 \Delta x_{it} + \Delta v_{it}$, and we assume $E[\Delta v_{it}|\Delta x_{it}] = 0$. Thus, degrees of freedom are reduced by 1 and equal to N (T - 1) - k.

FE: Approaches: (1) Demean the data (subtract mean of each variable for each city from its observations), (2) Include dummy variable for each city (city specific effects constant over time). $\Delta y_{it} = \beta_1 \Delta x_{it} + \Delta v_{it}$ (there is a ~ on top of each to indicate demean). FE estimators require a correction to the calculation of SEs. Because we demean, degrees of freedom are N(T-1) - k.

Serial Correlation occurs when e_{it} and $e_{i(t-1)}$ are correlated. Positive serial correlation is a problem, may underestimate SE estimates (negative correlation less problematic). In R, we use pgbtest(pmodel, h0="fd/fe", type=HC3) and pwfdtest(pmodel). If p-value <0.05, serial correlation is present H_0=no serial correlation. If idiosyncratic errors in levels are uncorrelated (FE hypothesis), first-differenced (FD) errors show serial correlation of 0.5. *If errors in levels follow a random walk (FD hypothesis), FD errors have no serial correlation. Both FE and FD are consistent, but FE is more efficient under "fe" and FD under "fd".* **Summary:** FD & FE & dummy inclusion control for time-constant unobserved factors, however no control over omitted variables varying over entities and time. Cannot include time-invariant factors (e.g. age), differencing will result in 0, but interaction possible. Missing data: FE handles seamlessly, FD loses two observations for each missing period, lowers sample size). Attrition occurs when units leave panel (they die, companies liquidate), if related to variables in model, leads to sample selection bias. FE can handle attrition correlated with a_i,(time-constant error) reducing impact.

Linear Probability Model (LPM) is so named because probability is linear in the parameters. P(approve= 1|x) = $\beta_0 + \beta_1$white. Where β_1 is the average, all else constant increase/decrease probability of success added to β_0. Advantage: coefficients represent change in probability for unit change in predictor, easy computation thanks to OLS, comparable to logistic regression when predictors are categorical. Disadvantage: predicted probabilities can fall outside of [0,1] range (nonsensical). Error variance depends on predicted probability, violating MLR5. Errors are not normally distributed (MLR6 violated). MLR5 -> Robust errors. MLR6 -> In LPMs, residuals are inherently not normal because y is bound to 0 and 1. Not critical if sample size is large, but for small sample inference it may be problematic. Robust SEs recommended. R^2 assumes continuous outcomes, so misleads interpretation. Instead include percent correctly predicted or naive benchmark (predicting majority class).
Interpretation: 4% to 6% is a 2% percentage point change but a 50% percent change.

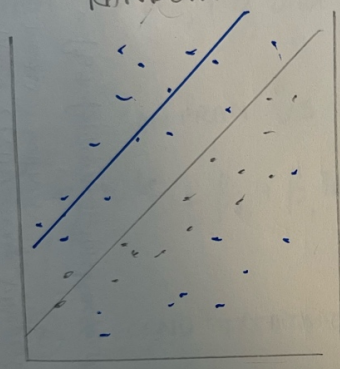
Non-linear (logistic & standard normal cumulative distribution): P(y= 1|x) = $G(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k) = G(\beta_0 + x\beta)$, where G is a cumulative distribution function. The difference between probit and logit models is small in terms of estimates and structure. Preferred over LPM because (1) Constrain predicted probabilities to lie between 0 and 1, (2) Ensure marginal effects are sensible across the range of explanatory variables. LPM provides simple and reasonable estimates and effective for statistical significance, however unsuitable for predicting individual probabilities.

CHEAT SHEET
(MLR4)

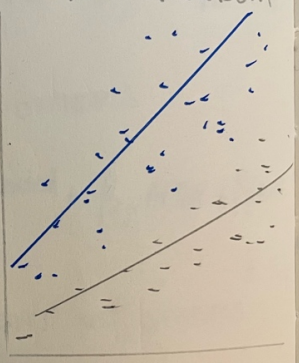
2.12.2024
[FSAR]

Measurement Error in y

RANDOM



NON RANDOM



"OBSERVED" (sample) "TRUE" (population)

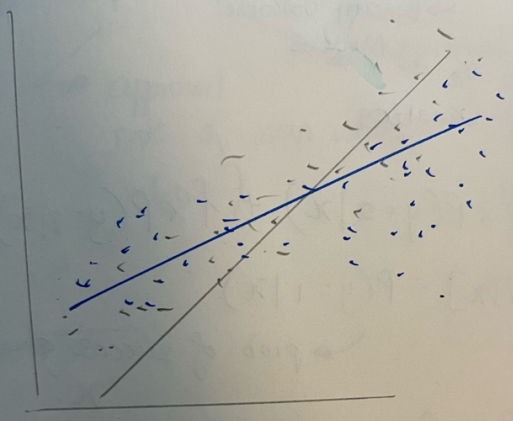
FORMULA

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u + e_0$$

e_0 ← measurement error (me)

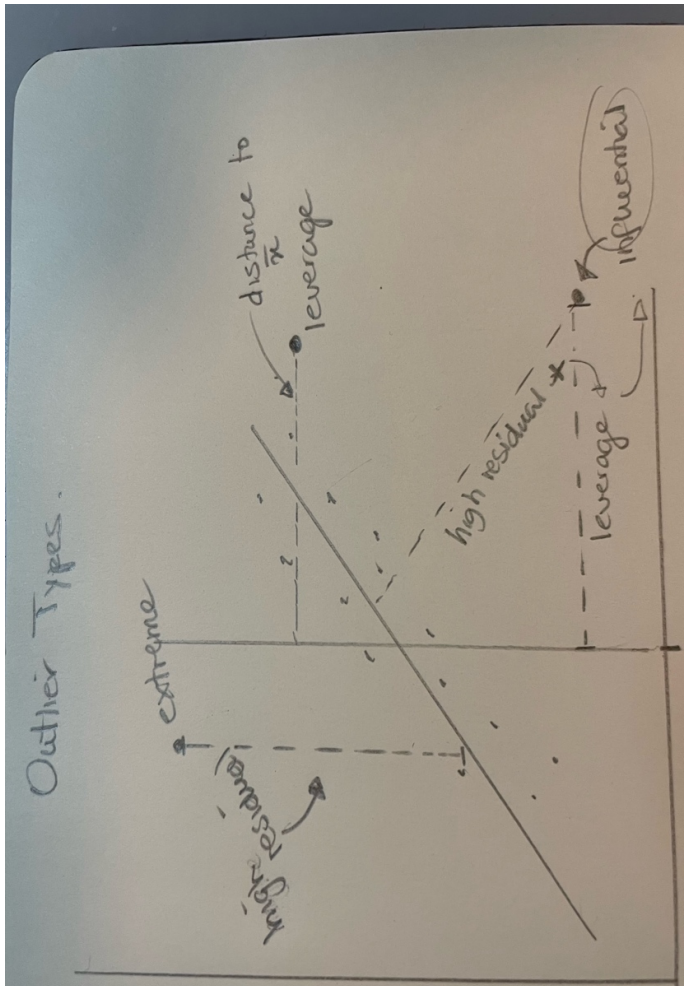
(true y) $y^* = y - e_0$

Measurement Error in x



"OBSERVED" (sample) "TRUE" (population)

Outlier Types.



PANEL DATA (OLS)

$HP_{it} = \beta_0 + \beta_1 crime_{it} + \nu_i + \alpha_t + u_{it}$ (trends in HP; $i = 1,\dots,N$, $t = 1,\dots,N$)

- β_1: magnitude of effect
- ν_i: city dependent (t-invariant) eg. geography, demographics, race
- α_t: time dependent
- u_{it}: idiosyncratic error (varies across i and t)

$HP_{it} = \beta_0 + \beta_1 crime_{it} + \gamma_1 \delta_{2t} + \gamma_2 \delta_{3t} + \dots + \gamma_{T-1} \delta_{Tt} + \alpha_t + w_{it}$ (dummies; 1st period is excluded)

$cov(\text{Actual variables}) \neq 0$ ⇒ Pooled ols is biased for this model ⇒ unobserved heterogeneity

PANEL DATA (FD) (NO INTERCEPT !!)

$HP_{it} - HP_{it-1} = \Delta HP_{it} = \beta_1 \Delta crime_{it} + \gamma_1 \Delta \delta_{2t} + (\alpha_t - \alpha_{t-1}) + \Delta u_{it}$ (α_t doesn't vary over time.)

Limitations:
1. not time-independent factors
2. small variation ⇒ high error

⇒ we need variance in t ⇒ CONSISTENT ESTIMATES ⇐ we need variance in i

⇐ assuming no heteroskedasticity
⇐ assuming no serial correlation

$cov(\Delta crime_{it}, \Delta u_{it}) = 0$

PANEL DATA (FE)

$HP_{it} = \beta_1 crime_{it} + \beta_2 unemploy_{it} + (\text{time dummies}) + (\alpha_i + u_{it})$ (composite error η_{it})

$\overline{HP}_i = \frac{1}{T}\sum_{t=1}^{T} HP_{it}$ → time averaged house price in given city.

$\overline{HP}_i = \beta_1 \overline{crime}_i + \beta_2 \overline{unemploy}_i + \alpha_i + \overline{u}_i$ (time constant as always the same)

$HP_{it} - \overline{HP}_i = \beta_1 (crime_{it} - \overline{crime}_i) + \beta_2 (unemploy_{it} - \overline{unemploy}_i) + \alpha_i - \alpha_i + (u_{it} - \overline{u}_i)$

$\left(cov(\tilde{x}_{it}, \tilde{u}_{it}) = 0\right)$

$\tilde{HP}_{it} = \beta_1 \tilde{crime}_{it} + \beta_2 \tilde{unemploy}_{it} + \tilde{u}_{it}$

Limitations
- removes time constants.

FE/FD vs Pooled OLS

$crime_{it} = \alpha_i + \beta\, unemploy + u_{it}$, we expect $\beta_0 \geq 0$ ✓

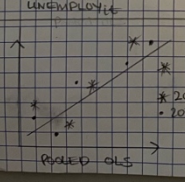
Pooled OLS (---)

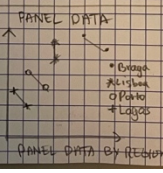
$\Delta crime_{it} = \beta_1 \Delta unemp_{it} + \Delta u_{it}$

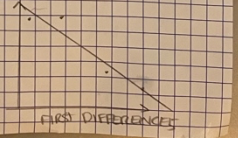
$\beta > 0$ ✓ FD (——)

$\beta < 0$ ✗ $\tilde{crime}_{it} = \tilde{unemploy}_{it} + \tilde{u}_{it}$ FE (——)

with $T = 2$, FD = FE







Interact dummy variables

$wage = \alpha + \beta_1 educ_i + \beta_2 sex_i + \beta_3 sex_i * educ_i$ (* interaction term)

$f = \alpha + \beta_1 educ + \beta_2(1) + \beta_3(1) * educ$

$= (\alpha + \beta_2) + (\beta_1 + \beta_3) educ$

$m = \alpha + \beta_1 educ$

- β_2: additional premium in wage (for being female)
- β_3: additional premium of having one more year of education for females.

Interact continuous - continuous

$Sales_i = \alpha + \beta_1 price + \beta_2 advert + \beta_3 price * advert_i$

$\beta_1 < 0$ $\quad \beta_2 > 0$ (assumption)

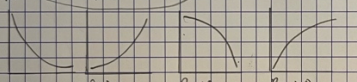
$A = 100$: $S_{A=100} = \alpha + \beta_1 price + \beta_2 (100) + (100)\beta_3 price$

$(\beta_1 + 100\beta_3)\, price$, $\beta_3 > 0$

"we expect β_3 to be positive where price sensitivity goes down the more we spend on advertisement"

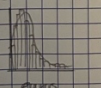
Quadratic terms

$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$ (β_2: rate of curvature; β_0: shift parameter (⊥ Δ))



general formula for turning point of any quadratic $x^* = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$

Logarithms (* not used for negative values)



level - level: $y = \beta_0 + \beta_1 x + u$

log - level: $\log(y) = \beta_0 + \beta_1 x + u$ ("for every percentage increase in x, y increase by...")

level - log: $y = \beta_0 + \beta_1 \log(x) + u$ ("for every x unit increase, y in %")

log - log: $\log(y) = \beta_0 + \beta_1 \log(x) + u$

DID regression (T=2)

$rprice = \beta_0 + \beta_1 y81 + \beta_2 nearinc + \beta_3 y81 * nearinc + u$ (CAUSAL IMPACT; AFTER * TREATED interaction)

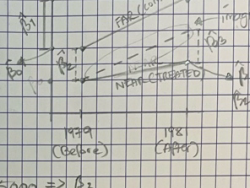
$\beta_0 = \bar{r}$ in 1978, far from incinerator

$\beta_0 + \beta_2 = \bar{r}$ in 1978, near incinerator

$\beta_0 + \beta_1 = \bar{r}$ in 1981, far from incinator

$\beta_0 + \beta_1 + \beta_2 + \beta_3 = \bar{r}$ in 1981, near incinerator

	Control	Treated	Diff
Before	β_0	$\beta_0 + \beta_2$	β_2
After	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_3$
Diff.	β_1	$\beta_1 + \beta_3$	β_3



T	Before (1979)	T	After (1981)
0	100000	0	150000
1	70000	1	85000

(1979 (Before), 1981 (After))

$(150000 - 100000) - (85000 - 70000) = 50000 \Rightarrow \beta_3$

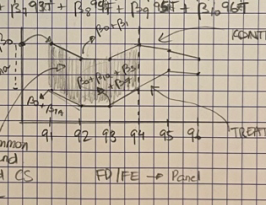
① Manual: $(\bar{Y}_{TA} - \bar{Y}_{TB}) - (\bar{Y}_{CA} - \bar{Y}_{CB})$

② Graphical:

DID regression (T > 2)

work: $\beta_0 + \beta_1 y2 + \beta_2 y3 + \beta_3 y4 + \beta_4 y5 + \beta_5 y6 + \beta_6 y7 + \beta_7 y8 + \beta_8 y9 + \beta_9 y10 + \beta_{10} y11 + \beta_{11} nearinc + \beta_{12} y2 * nearinc + \dots + \beta_{21} y11 * nearinc$

Year	Control	Treated	Diff
91	β_0	$\beta_0 + \beta_{11}$	β_{11}
92	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_{11} + \beta_{12}$	$\beta_{11} + \beta_{12}$
93	$\beta_0 + \beta_2$	$\beta_0 + \beta_2 + \beta_{11} + \beta_{13}$	$\beta_{11} + \beta_{13}$
94	$\beta_0 + \beta_3$	$\beta_0 + \beta_3 + \beta_{11} + \beta_{14}$	$\beta_{11} + \beta_{14}$
95	$\beta_0 + \beta_4$	$\beta_0 + \beta_4 + \beta_{11} + \beta_{15}$	$\beta_{11} + \beta_{15}$
96	$\beta_0 + \beta_5$	$\beta_0 + \beta_5 + \beta_{11} + \beta_{16}$	$\beta_{11} + \beta_{16}$



OLS → Cross-section → Pooled CS; Panel

Diff in diff → Pooled CS; panel

FD/FE → Panel

Linear Probability Model (LPM)

$approve = \beta_0 + \beta_1 White + \varepsilon$

$P(y = 1|x) = E[y|x]$

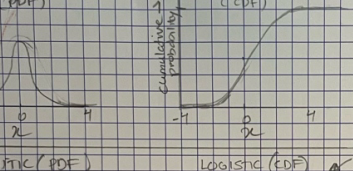
probability of success = expected value

$\therefore P(approve = 1|x) = \beta_0 + \beta_1 White$

Non-linear (probit model)

We assume a standard normal cumulative distribution for $G(\beta_0 + x\beta)$

$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(z)\, dz$, where $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ ⇒ pdf of standard normal distribution

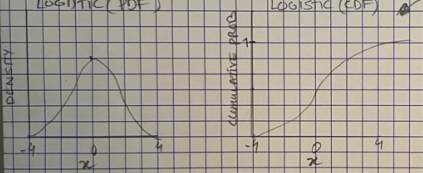


Non-linear (logit)

II - II (same assumption) but

$G(z) = \Lambda(z) = \frac{e^z}{1+e^z}$

ensures that $P(y=1|x)$ is between 0 and 1 for all values of parameters and the x_j



PARTIAL EFFECTS

continuous x_j:

$\frac{\partial G(\beta_0 + x\beta)}{\partial x_j} = g(\beta_0 + x\beta)\beta_j$

$g(z) = \frac{\partial G(z)}{\partial z}$, $g(z) = PDF$

discrete x_j (Let $\pi = P(y=1|x) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 \log(x_2) + \beta_4 x_3)$)

- $\frac{\partial G(\beta_0 + x\beta)}{\partial x_1} = g(\beta_0 + x\beta)(\beta_1 + 2\beta_2 x_1)$
- $\frac{\partial G(\beta_0 + x\beta)}{\partial x_2} = g(\beta_0 + x\beta)\left(\frac{\beta_3}{x_2}\right)$
- $\frac{\partial G(\beta_0 + x\beta)}{\partial x_3} = g(\beta_0 + x\beta)\beta_4$

$\beta_0 + x\beta$ ← scale factor; depends on z (all explanatory variables)

PEA (partial effect at average) → $g(\hat{\beta}_0 + \bar{z}\hat{\beta})$

then $g(\hat{\beta}_0 + \bar{z}\hat{\beta})\beta_4$ → PEA or MEM (marginal effect at the mean)

APE or MNE (average partial effect / average marginal effect)

$\therefore \frac{1}{N}\sum_{i=1}^{n} g(\hat{\beta}_0 + x_i\hat{\beta}) \therefore \rightarrow \left[\frac{1}{N}\sum_{i=1}^{n} g(\hat{\beta}_0 + x_i\hat{\beta})\right]\hat{\beta}_j$

Interact dummy variables

$wage = \alpha + \beta_1 educ_i + \beta_2 sex_i + \boxed{\beta_3} sex_i educ_i$ ← interaction term

$f = \alpha + \beta_1 educ + \beta_2(1) + \beta_3(1) * educ$

$= (\alpha + \beta_2) + (\beta_1 + \beta_3) educ$

$m = \alpha + \beta_1 educ$

- additional premium in wage (for being female)
- additional premium of having one more year of education for females.

Interact continuous - continuous

$Sales_i = \alpha + \beta_1 price + \beta_2 advert + \beta_3 price * advert_i$

$\beta_1 < 0 \quad \beta_2 > 0$ (assumptions)

$A = 100$:

$S_{A=100} = \alpha + \beta_1 price + \beta_2(100) + (100)\beta_3 price$

$(\beta_1 + 100\beta_3)\, price$, $\beta_3 > 0$

"we expect β_3 to be positive where price sensitivity goes down the more we spend on advertisement"

Quadratic terms

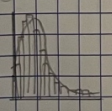
$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$ — β_2: rate of curvature; β_1: shift parameter ($\lessgtr 0$)

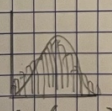
$\beta_1 < 0$	$\beta_1 > 0$	$\beta_1 < 0$	$\beta_1 > 0$
$\beta_2 > 0$	$\beta_2 > 0$	$\beta_2 < 0$	$\beta_2 < 0$

general formula for turning point of any quadratic: $x^* = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$

Logarithms

* not used for negative values





level - level	level - log	log - level	log - log
$y = \beta_0 + \beta_1 x + u$	$y = \beta_0 + \beta_1 \log(x) + u$	$\log(y) = \beta_0 + \beta_1 x + u$	$\log(y) = \beta_0 + \beta_1 \log(x) + u$
	"For every 1% increase in x, y increases by ..."	"For every percentage increase in x, y increase by ..."	

PANEL DATA (OLS)

$HP_{it} = \beta_0 + \beta_1 crime_{it} + v_t + \alpha_i + u_{it}$

- β_1: magnitude of effect
- v_t: time dependent (trends in HP)
- α_i: city dependent (time constant) eg. geography, demographics, race
- u_{it}: idiosyncratic error (varies across i and t)

$i = 1, \dots, N$; $t = 1, \dots, N$

$HP_{it} = \beta_0 + \beta_1 crime_{it} + \gamma_1 \delta_{1t} + \gamma_2 \delta_{2t} + \dots + \gamma_{T-1} \delta_{Tt} + \alpha_i + u_{it}$

(dummies) — 1st period is excluded. $\alpha_i + u_{it} = \eta_{it}$

$cov(x_{it}, \eta_{it}) \neq 0$ ∴ Pooled OLS is biased for this model ← unobserved heterogeneity

PANEL DATA (FD)

* NO INTERCEPT !!!

$HP_{it} - HP_{it-1} = \Delta HP_{it} = \beta_1 \Delta crime_{it} + \gamma_1 \Delta\delta_{2t} + (\alpha_i - \alpha_i) + \Delta u_{it}$ (α_i doesn't vary over time)

Limitations:

1) time independent factors: ↳ we need variance in t ⇒ CONSISTENT ESTIMATES ⇐ ↳ we need variance in i
2) small variation → high error

$cov(\Delta crime_{it}, \Delta u_{it}) = 0$ ↳ assuming no heteroskedasticity ↳ assuming no serial correlation

PANEL DATA (FE)

$HP_{it} = \beta_1 crime + \beta_2 unemploy_{it} + (time\ dummies) + \alpha_i + u_{it}$

composite error η_{it}

$\overline{HP_i} = \frac{1}{T}\sum_{t=1}^{T} HP_{it}$ → time averaged house price in given city.

$\overline{HP_i} = \beta_1 \overline{crime_i} + \beta_2 \overline{unemploy_i} + \alpha_i + \overline{u_i}$ (time constant so stays the same)

$HP_{it} - \overline{HP_i} = \beta_1 (crime_{it} - \overline{crime_i}) + \beta_2 (unemploy_{it} - \overline{unemploy_i}) + \alpha_i - \alpha_i + (u_{it} - \overline{u_i})$

$(cov(x_{it}, u_{it}) = 0)$

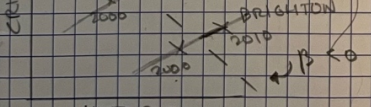
$\widetilde{HP}_{it} = \beta_1 \widetilde{crime}_{it} + \beta_2 \widetilde{unemploy}_{it} + \widetilde{u}_{it}$

Limitations
↳ removes time constants.

FE/FD vs Pooled OLS

$crime_{it} = \alpha_i + \beta unemploy + u_{it}$, we expect $\beta_0 > 0$ ✗

Pooled OLS (---)

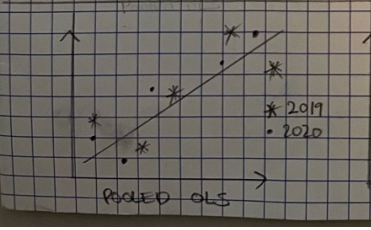


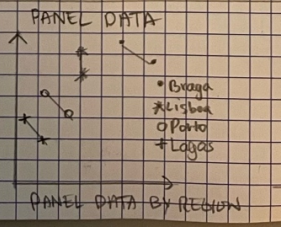
$\Delta Crime_{it} = \beta_2 \Delta unemp_{it} + \Delta u_{it}$

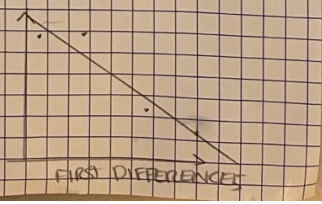
$\beta > 0$ ✓ FD (——)

$\beta < 0$ ✗ $\widetilde{crime}_{it} = \widetilde{unemploy}_{it} + \widetilde{u}_{it}$ FE (——)

with $T = 2$, FD = FE







Linear Probability Model (LPM)

approve = $\beta_0 + \beta_1$ white + ε

$P(y=1|x) = E[y|x]$

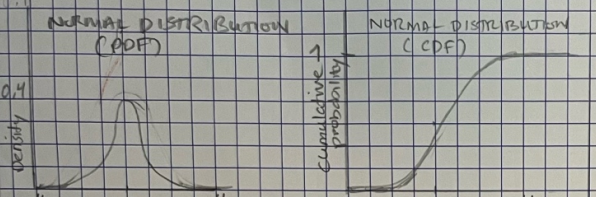
probability of success = expected value

$\therefore P(\text{approv}=1|x) = \beta_0 + \beta_1 \text{white}$

Non-linear (probit model)

We assume a standard normal cumulative distribution for $G(\beta_0 + x\beta)$

$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(z)\,dz$, where $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ ⇒ pdf of standard normal distribution

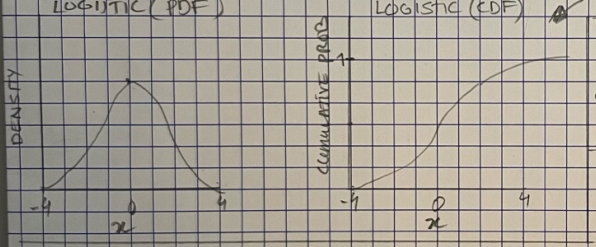


Non-linear (logit)

I + II (same assumption) but

$G(z) = \Lambda(z) = \frac{e^z}{1+e^z}$

ensures that $P(y=1|x)$ is between 0 and 1 for all values of parameters and the x_j



PARTIAL EFFECTS

continuous x_j:

$\frac{\partial G(\beta_0 + x\beta)}{\partial x_j} = g(\beta_0 + x\beta)\beta_j$

$g(z) = \frac{\partial G(z)}{\partial z}$, $g(z)$ = PDF

discrete x_j: Let $z = P(y=1|x) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 \log(x_2) + \beta_4 x_2)$

- $\frac{\partial G(\beta_0 + x\beta)}{\partial x_1} = g(\beta_0 + x\beta)(\beta_1 + 2\beta_2 x_1)$
- $\frac{\partial G(\beta_0 + x\beta)}{\partial x_2} = g(\beta_0 + x\beta)\left(\frac{\beta_3}{x_2}\right)$
- $\frac{\partial G(\beta_0 + x\beta)}{\partial x_3} = g(\beta_0 + x\beta)\beta_4$

$\beta_0 + x\beta$ ← scale factor

⇕ depends on z (all explanatory variables)

PEA (partial effect at average) → $g(\hat{\beta}_0 + \bar{z}\hat{\beta})$

then $g(\hat{\beta}_0 + \bar{z}\hat{\beta})\beta_4$ → PEA or MEM (marginal effect at the mean)

APE or AME (average partial effect) / (average marginal effect)

$\therefore \frac{1}{N}\sum_{i=1}^{n} g(\hat{\beta}_0 + x_i\hat{\beta}) \therefore \rightarrow \left[\frac{1}{N}\sum_{i=1}^{n} g(\hat{\beta}_0 + x_i\hat{\beta})\right]\hat{\beta}_j$

DID regression (T=2)

rprice = $\beta_0 + \beta_1 y81 + \beta_2 nearinc + (\beta_3 y81 * nearinc) + u$ → CAUSAL IMPACT

$y81 * nearinc$ → AFTER * TREATED interaction

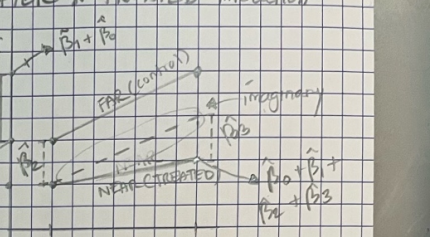
$\beta_0 = \bar{r}$ in 1978, far from incinerator

$\beta_0 + \beta_2 = \bar{r}$ in 1978, near incinerator

$\beta_0 + \beta_1 = \bar{r}$ in 1981, far from incinerator

$\beta_0 + \beta_1 + \beta_2 + \beta_3 = \bar{r}$ in 1981, near incinerator

	Control	Treated	Diff
Before	β_0	$\beta_0 + \beta_2$	β_2
After	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_3$
Diff.	β_1	$\beta_1 + \beta_3$	β_3

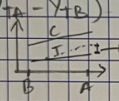


T	Before (1979)	T	After (1981)
0	100000	0	120000
1	70000	1	85000

$(120000 - 100000) - (85000 - 70000) = 5000 \Rightarrow \beta_3$

① Manual : $(\bar{Y}_{TA} - \bar{Y}_{TB}) - (\bar{Y}_{CA} - \bar{Y}_{CB})$

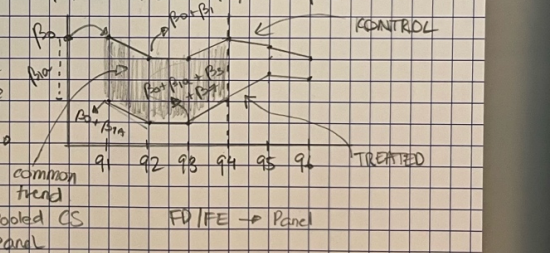
② Graphical :



DID regression (T>2)

work = $\beta_0 + \beta_1 92 + \beta_2 93 + \beta_3 94 + \beta_4 95 + \beta_5 96 + \beta_6 92T + \beta_7 93T + \beta_8 94T + \beta_9 95T + \beta_{10} 96T$

Year	Control	Treated	Diff
91	β_0	$\beta_0 + \beta_{1a}$	β_1
92	$\beta_0 + \beta_1$	$\beta_0 + \beta_{1a} + \beta_1 + \beta_6$	$\beta_{1a} + \beta_6$
93	$\beta_0 + \beta_2$	$\beta_0 + \beta_{1a} + \beta_2 + \beta_7$	$\beta_{1a} + \beta_7$
94	$\beta_0 + \beta_3$	$\beta_0 + \beta_{1a} + \beta_3 + \beta_8$	$\beta_{1a} + \beta_8$
95	$\beta_0 + \beta_4$	$\beta_0 + \beta_{1a} + \beta_4 + \beta_9$	$\beta_{1a} + \beta_9$
96	$\beta_0 + \beta_5$	$\beta_0 + \beta_{1a} + \beta_5 + \beta_{10}$	$\beta_{1a} + \beta_{10}$



OLS → Cross-section / Pooled CS / Panel

Diff. in diff → Pooled CS / Panel

FD/FE → Panel