MICROECONOMETRICS II

2020/21

Framework

Course description:

- Addresses complementary topics relative to Microeconometrics I
- Foccus on cross-sectional and panel data:
  - The interaction between theory and empirical econometric analysis is emphasized
  - Students will be trained in formulating and testing economic models using real data

Pre-requisites: Econometrics, Microeconometrics I

Basic reference: Cameron A.C. e Trivedi P.K. (2005), Microeconometrics, Methods and Applications, Cambridge University Press.

Framework

List of topics:

- Estimation and inference in nonlinear models: extremum estimators, GMM estimators, maximum likelihood estimators. Specification Analysis. Panel Data
- Models for count data or positive dependent variable
- Models for fractional data
- Models for truncated, cesored, excess of zeros, or with endogenous stratification data
- Models for duration data
- Quantile regression

Framework

Assessment:

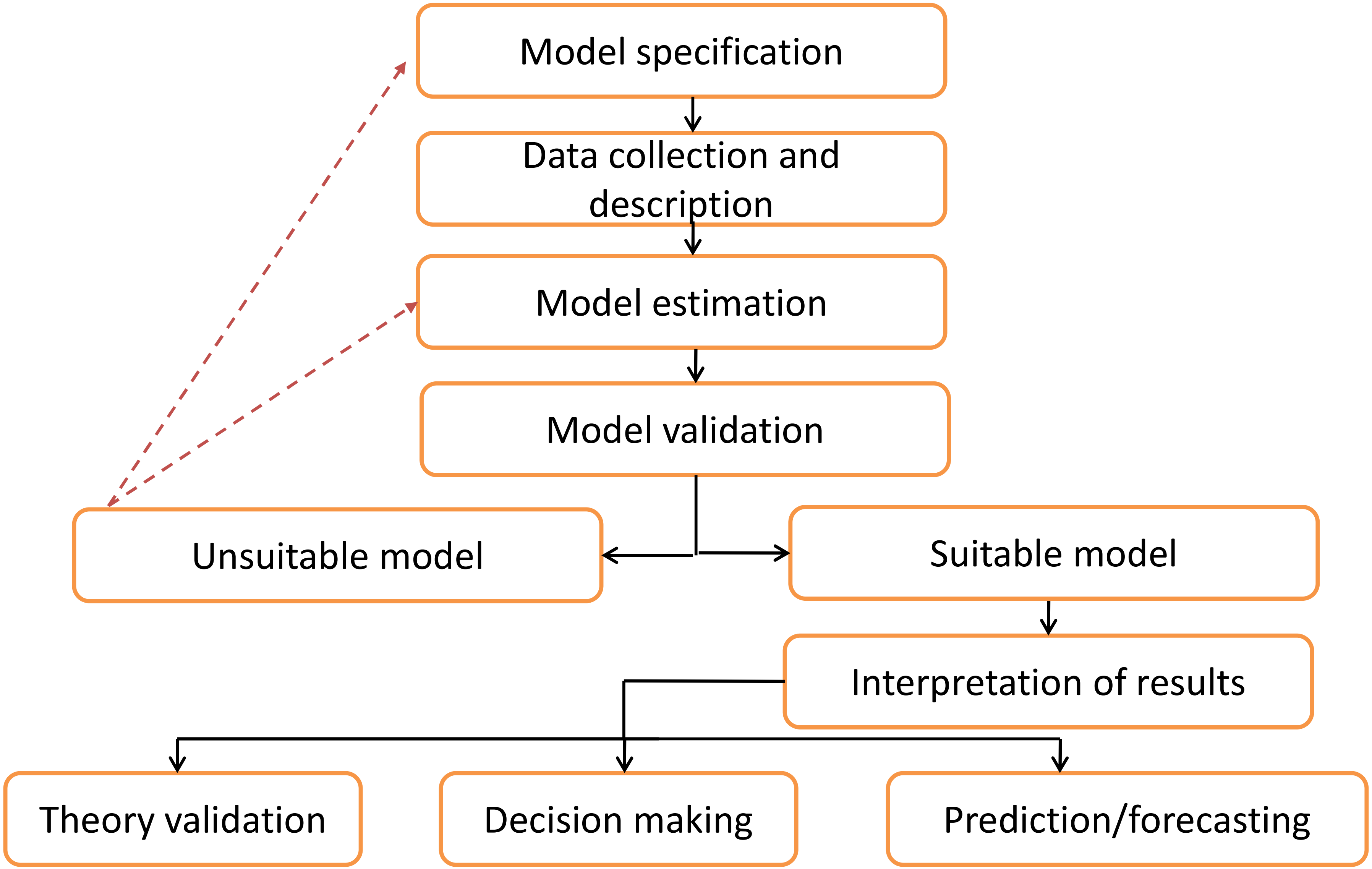
“Época Normal”: empirical project (groups of 3 / 4 students) (50%) + open book test (50%). Minimum classification in the test: 7.

“Época de Recurso”: open book test (100%).

Empirical project:

- 26 October: preliminary presentation (data description, ideas for methodology, …)
- 14 December: final presentation
- 16 December: delivery of the final version

Methodology



Main types of microeconometric models

- Regression Model: explaining $E(Y|X)$

- Probabilistic model:
  - Aim: Explaining $P(Y|X)$
  - Usually incorporates a regression model for $E(Y|X)$ e $P(Y|X)$

Y: dependent variable

X: explanatory variables

$E(Y|X)$: expected value for Y given X

$P(Y|X)$: probability of Y being equal to a specific value given X

Dependent Variable versus Econometric Model

The characteristics of the dependent variable restricts the models that may be applied in each case:

Y	**Type of outcome**	**Main model**
$]-\infty,+\infty[$	Unbounded data	Linear
$\{0,1\}$	Binary choices	Logit,...
$\{0,1,2,\dots,J-1\}$	Multinomial choices	Multinomial logit,...
$\{0,1,2,\dots,J-1\}$	Ordered choices	Ordered logit,...
$\{0,1,2,\dots\}$	Count data	Poisson,...
$[0,+\infty[$	Nonnegative data	Exponential
$[0,1]$	Fractional data	Fractional Logit,...

Quantities of interes in nonlinear models

- Partial effects over
  - $E(Y|X) = G(X\beta)$
  - $P(Y|X) = F(X\beta)$
  - Effects: $\Delta X_j = 1 \Longrightarrow$
  - $\Delta E(Y|X) = \dfrac{\partial E(Y|X)}{\partial X_j} = \dfrac{\partial G(X\beta)}{\partial X_j} = \beta_j \dfrac{\partial G(X\beta)}{\partial X\beta} = \beta_j g(x_i'\beta)$
  - $\Delta P(Y|X) = \dfrac{\partial P\ \ (Y|X)}{\partial X_j} = \dfrac{\partial F(X\beta)}{\partial X_j} = \beta_j \dfrac{\partial F(X\beta)}{\partial X\beta} = \beta_j f(x_i'\beta)$

  * Adaptation for discrete X_j

- Other quantities of interest
  - Example: when modelling a nonnegative outcome, $Y \geq 0$ with lots of zeros, it may be interesting to estimate also:
    - » $P(Y = 0|X)$
    - » $E(Y|X, Y > 0)$

Quantities of interes in nonlinear models

- Partial effects may be compared across different models, but the values of β cannot, in general

- However, because $\frac{\partial G(X\beta)}{\partial X\beta} > 0$ and $\frac{\partial F(X\beta)}{\partial X\beta} > 0$:
  - The sign of the partial effect is given by the sign of β_j
  - Testing the statistical significance of the partial effect is equivalent to test for $H_0: \beta_j = 0$

- Calculation of the magnitude of the partial effects:
  - Calculate the partial effects for each individual in the sample and then obtain the mean of those effects
  - Replace x by its sample means
  - Replace x by specific values

Stata
(after estimating the model)

margins, dydx(*varlist*)

margins, dydx(*varlist*) atmeans

margins, dydx(*varlist*) at(...)

Model Transformations and Adaptations

Bounded continuous outcomes may be transformed in such a way that they give rise to unbounded outcomes which may be modelled using a linear model

- this requires retransformation back to the original scale

Any microeconometric model may require adaptations:

- Data structure: cross-section, panel
- Non-random samples: stratified, censored, truncated
- Measurement error
- Endogenous explanatory variables
- Corner solutions / excess of zeros

Types of Explanatory Variables

Quantitative variables:

- Levels (Euro, kilograms, meters,...)
- Levels and squares
- Logs
- Growth rates
- Per capita values

Qualitative variables

- Binary (dummy) variables:

$$X = \{0,1\}$$

- Interaction variables:

$$X = Dummy\ var. * (Quantitative\ or\ dummy\ var.)$$

Extremum estimators

An extremum estimator $\hat{\theta}$ maximizes or minimizes a given objective function, defined as a function of a sample of i=1,...n individuals with data on $Z = (Y, X)$ and the k-dimensioned vector of parameters θ:

$$\theta = \operatorname*{argmax}_{\theta \in \Theta} Q_n(\theta)$$

The estimator $\hat{\theta}$ is the solution of the first order condition

$$\nabla_\theta Q_n(\theta)\Big|_{\theta=\hat{\theta}} = 0$$

Extremum estimators

Particular cases:

1. M Estimators: maximum likelihood like

$$Q_n(\theta) = \frac{1}{n}\sum\nolimits_{i=1}^{n} q(z_i, \theta)$$

2. GMM: minimize a quadratic form of averages of moment conditions

$$Q_n(\theta) = -g_n \widehat{W} g_n$$

where $g_n = \frac{1}{n}\sum_{i=1}^{n} g(z_i, \theta)$ reflect s, $s \geq k$, moment conditions $\mathrm{E}[g(z, \theta_0,)] = 0$ and W is a weigthing matrix

- W=I in just-identified cases (s=k)

Extremum estimators

3. OLS estimator: $Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}[y_i - x\theta]^2$

4. NLS estimator: $Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}[y_i - h(x,\theta)]^2$, where $h(.)$ is a nonlinear function

5. ML estimator: $Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} lnf(x_i,\theta)$, where $f(.)$ is the pdf assumed

OLS, NLS, and ML estimators are particular cases of both M and GMM estimators, with W=I and first order conditions used as moment conditions

Extremum estimators

6. LAD (least absolute deviation) estimator

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} |y_i - med(Y|X)|$$

7. Manski's maximum score estimator (binary models)

$$Q_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} |y_i - 1(x\theta > 0)|$$

where $1(.)$ is an indicator function

GMM estimators

According to Newey and McFadden (1994)

1. Consistency - Theorem 2.6 - under the established regularity conditions, the GMM estimator $\hat{\theta}$ converges to the true value θ_0

$$plim\hat{\theta} = \theta_0$$

2. Asymptotic normality - Theorem 3.4

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \overset{d}{\rightarrow} N[0, (A'WA)^{-1}A'WBWA(A'WA)^{-1}]$$

where $\mathrm{A} = \mathrm{E}[\nabla_\theta g(z, \theta_0)]$ and $\mathrm{B} = \mathrm{E}[g(z, \theta_0) g(z, \theta_0)']$)

GMM estimators

3. Asymptotic efficiency - Theorem 5.2 – assuming B as a nonsingular matrix, the GMM estimator with $\mathrm{W} = \mathrm{plim}\widehat{W} = B^{-1}$ is asymptotically efficient within the class of GMM estimators.

- In this case, the following simplification is obtained

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N[0, (A'B^{-1}A)^{-1}]$$

- If, additionally, we have exact identification (s=k: same number of parameters to estimate and moment conditions):

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N[0, A^{-1}BA'^{-1}]$$

- Estimator of the covariance matrix $\hat{C}_n$: replace A and B by A_n and B_n, which result from the replacement of $E(.)$ by $\frac{1}{n}\sum_{i=1}^{n}$.

GMM estimators

Alternative forms of the weighting matrix W:

Different GMM estimators result from different forms of W

- **Two-step**:
  - Use $W = I$ and estimate $\hat{\theta}_{1step}$
  - Use $\hat{\theta}_{1step}$ in W and then estimate $\hat{\theta}_{2step}$
- **Iterated**: repeats two-step ($\hat{\theta}_{2step}$ is used to compute W of a 3rd step, $\hat{\theta}_{3step}$ is used to compute the W of a 4rd step, …)
- **Continuously updating GMM**: in each step, $\hat{\theta}$ is estimated from both from the moment conditions and W, which is designated as $W(\theta)$

ML and QML estimators

Review…

Requirement: specification of

- The G function in $E(Y|X) = G(X\beta)$
- The F function in $Pr(Y|X) = F(X\beta)$

Main assumptions:

- ML:
  - Correct specification of both G and F
- QML:
  - Correct specification of G
  - F does not need to be correctly specified but needs to be included in the linear exponential family (e.g. Normal, Bernoulli, Poisson, Exponencial, Gama, etc.)

ML and QML estimators

ML / QML estimation:

- Density function: $f(y_i|x_i;\theta)$
- Likelihood function:
  - Gives the probability for the occurence of a full set of sample values on the assumption that the density function $f(y_i|x_i;\theta)$ is correct
  - Assuming independence, it is calculated as

$$L = \prod_{i=1}^{n} f(y_i|x_i;\theta)$$

ML and QML estimators

- The parameter θ is unknown and its value is chosen in order to maximize L and thus to maximize the probability that the sample values were in fact generated by the chosen density $f(\cdot)$:

$$max_{\theta} L = \prod_{i=1}^{n} f(y_i|x_i;\theta)$$

- Actually, it is more common to maximize $LL = ln(L)$:

$$max_{\theta} LL = \sum_{i=1}^{n} ln[f(y_i|x_i;\theta)]$$

  - It is easier to maximize
  - It produces the sames estimates for θ

ML and QML estimators

Quantities to take into account:

- Score vector: $\mathrm{S}(\theta) = \nabla_\theta ln[f(y|\, x; \theta)]$
- Hessean matrix: $\mathrm{H}(\theta) = \nabla_{\theta\theta'} ln[f(y|x;\theta)]$
- Symmetric of the expected hessean matrix: $\mathrm{A} = \mathrm{E}[-\mathrm{H}(\theta)]$
- Fisher information matrix: $\mathcal{J} = \mathrm{B} = \mathrm{E}[S(\theta)S(\theta)']$
- Information matrix equality: $\mathcal{J} = \mathrm{E}[-\mathrm{H}(\theta)]$ or B=A

ML and QML estimators

Properties of ML estimators:

- Consistency - Theorem 2.5: $plim\hat{\theta} = \theta_0$

- Assymptotic normality - Theorem 3.3: $\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N[0, \mathcal{I}^{-1}]$, which results from $\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N[0, A^{-1}BA'^{-1}]$ with B=A

- Assymptotic efficienc - Theorem 5.1: this theorem refers to GMM, but establishes $\mathcal{I}^{-1}$ as the inferior limit of the GMM covariace matrix, which is the ML covariace matrix.

ML and QML estimators

Properties of QML estimators:

- Consistency : $plim\hat{\theta} = \theta_0$
- Assymptotic normality: $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N[0, A^{-1}BA'^{-1}]$ as the information matrix equality does not hold (GMM estimator with exact identification)
- Assymptotic efficiency requires the additional assumption of correct specification of $V(Y|X)$. In these conditions $A = B$ and $A^{-1}BA'^{-1}$ simplifies to the ML result $\mathcal{I}^{-1}$

* In ML estimators, the designated robust covariance matrix $A^{-1}BA'^{-1}$ may be used instead of the standard form $\mathcal{I}^{-1}$. Additional alternatives: cluster-robust (panel data) and boostrap

GMM hypothesis testing

1. Wald, LM, and LR tests

Define r restrictons $r(\theta) = 0$ with $R(\theta) = \nabla_\theta r(\theta)$ and denote the estimators of the restricted and unrestricted model as θ_r and θ_u, respectively

Theorem 9.2: all the following test statistics converge in distribution to χ_r^2:

$$W = nr(\hat{\theta}_u)'[\hat{R}\hat{C}^{-1}\hat{R}']r(\hat{\theta}_u)$$

$$LM = n\nabla_\theta Q_n(\hat{\theta}_r)'\hat{B}^{-1}\hat{A}\hat{C}\hat{A}'\hat{B}^{-1}\nabla_\theta Q_n(\hat{\theta}_r)$$

$$LR = n[Q_n(\hat{\theta}_r) - Q_n(\hat{\theta}_u)]$$

where $\hat{C}$ is the covariance matrix estimator

GMM hypothesis testing

Particular case: ML

Test for the joint significance of a set of parameters:

- Competing models:
  - Restricted (smaller) model, based on $L_R(\beta_0 + \beta_1 x_1 + \cdots + \beta_g x_g)$
  - Full (larger) model , based on $L_F(\beta_0 + \beta_1 x_1 + \cdots + \beta_g x_g + \beta_{g+1} x_{g+1} + \cdots + \beta_k x_k)$

- Hypotheses:

  H_0: $\beta_{g+1} = \cdots = \beta_k = 0$ (restricted model)

  H_1: No H_0 (full model)

GMM hypothesis testing

- LR test:

$$LR = 2[LL_F(X\beta_F) - LL_R(X\beta_R)] \sim \chi^2_{k-g}$$

  - Available in most econometric packages
  - Easy calculation
  - Both the competing models need to be estimated

Stata
(estimate one model)
estimates store *Model1*
(estimate the other model)
estimates store *Model2*
lrtest *Model*1 *Model*2

- Wald test:

$$W = \hat{\beta}_D' \left[\mathrm{Var}\left(\hat{\beta}_D\right)\right]^{-1} \hat{\beta}_D \sim \chi^2_{k-g}$$

where $\hat{\beta}_D = \left(\hat{\beta}_{g+1}, \dots, \hat{\beta}_k\right)$ is estimated based on $LL_F(X\beta_F)$

  - When H_0: $\beta_j = 0$, W simplifies to:

$$t = \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim \mathcal{N}(0,1)$$

  - Available in most econometric packages
  - Only the full model needs to be estimated

Stata
(after estimating the full model)
test $X_{g+1} \cdots X_k$

GMM hypothesis testing

- Score/LM test:

$$\text{Score} = \frac{\partial LL_F\left(X\hat{\beta}_M\right)}{\partial \beta}\left[Var_F\left(\hat{\beta}_M\right)\right]^{-1}\frac{\partial LL_F\left(X\hat{\beta}_M\right)}{\partial \beta} \sim \chi^2_{k-g}$$

where $\hat{\beta}_M = \left(\hat{\beta}_0, \dots, \hat{\beta}_g, 0, \dots 0\right)$, with $\left(\hat{\beta}_0, \dots, \hat{\beta}_g\right)$ estimated based on $LL_R(X\beta_R)$

  - Only the restricted model needs to be estimated, which may be an advantage when the full model is complex and hard to estimate
  - Rarely available in econometric packages, requiring programming

GMM hypothesis testing

2. Overidentification tests

The J test of Hansen (1982) checks whether all the moment conditions are statisfied in the data

$H_0 \colon \mathrm{E}[g(z,\theta)] = 0$
$H_1 \colon \mathrm{E}[g(z,\theta)] \neq 0$

$$J = nQ_n(\hat{\theta}) \sim \chi^2_{s-k}$$

GMM hypothesis testing

2. Overidentification tests

Eichenbaum, Hansen e Singleton (1988) proposed an extension to check the validity of a subset of moment conditions. Partition the vector of moment conditions as $g(z,\theta)' = [g_1(z,\theta)', g_2(z,\theta)']$ and suppose that the aim is testing the validity of the last s_2 moment conditions:

$H_0\colon \mathrm{E}[g_2(z,\theta)] = 0$

$H_1\colon \mathrm{E}[g_2(z,\theta)] \neq 0$

In addition to θ also θ_1 is estimated (from $g_1(z,\theta)$) and the test statistics is

$$J_2 = n\left[Q_n(\hat{\theta}) - Q_n(\hat{\theta}_1)\right] \sim \chi^2_{s_2}$$

GMM hypothesis testing

3. Conditional moment tests

Independentemently proposed by Newey (1985) and Tauchen (1985), checks wheter a subset of moment conditions is satisfied, similarly to J_2:

H_0: $\mathrm{E}[g_2(z,\theta)] = 0$

H_1: $\mathrm{E}[g_2(z,\theta)] \neq 0$

However, in this framework only the restricted model, based on $g_1(z,\theta)$, is estimated. The test statistic is

$$CM = n\hat{g}'_{2n}\left(\hat{\theta}_1\right)\left(\hat{D}_n\hat{B}_n\hat{D}'_n\right)\hat{g}_{2n}\left(\hat{\theta}_1\right) \sim \chi^2_{(s_2)}$$

where $D = -A_2\left(A'_1 B_{11}^{-1} A_1\right)^{-1} A'_1 B_{11}^{-1} I_{s2}$

GMM hypothesis testing

4. Hausman test

Consider two estimators $\hat{\theta}$ and $\tilde{\theta}$. Under H_0 both are consistent but $\tilde{\theta}$ is more efficient. Under H_1 only $\hat{\theta}$ is consistent.

H_0: $\text{plim}\left(\hat{\theta} - \tilde{\theta}\right) = 0$

H_1: $\text{plim}\left(\hat{\theta} - \tilde{\theta}\right) \neq 0$

$$H = \frac{1}{n}\left(\hat{\theta} - \tilde{\theta}\right)' \left[V\left(\hat{\theta}\right) - V\left(\tilde{\theta}\right)\right]\left(\hat{\theta} - \tilde{\theta}\right) \sim \chi^2_{rank\left[V\left(\hat{\theta}\right) - V\left(\tilde{\theta}\right)\right]}$$

Specification testing

Specification tests:

- For $E(Y|X)$:
  - RESET test
  - Chow test

- For $Pr(Y|X)$:
  - Information Matrix text, usually very hard to implement
  - More common: tests designed specifically to particular models

Specification testing

RESET test:

- Implementation:
  - Estimate the original model:

$$Pr(Y|X) = F(\beta_0 + \beta_1 x_1 + \cdots \beta_k x_k)$$

  - Generate the variables $(X\hat{\beta})^2, (X\hat{\beta})^3, (X\hat{\beta})^4, \ldots$
  - Add the generated variables to the original model and estimate the following auxiliary model:

$$Pr(Y|X) = F\left[\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma_1 (X\hat{\beta})^2 + \gamma_2 (X\hat{\beta})^3 + \gamma_3 (X\hat{\beta})^4 + \cdots\right]$$

  - Apply a LR / Wald test for the significance of the added variables:

    $H_0: \gamma_1 = \gamma_2 = \gamma_3 = \cdots = 0$ (suitable model functional form)

    H_1: No H_0 (unsuitable model functional form)

Specification testing

Chow Test for Structural Breaks:

- Context:
  - Two groups of individuals / firms / ...: G_A, G_B
  - It is suspected that the behaviour of the two groups in which regards the dependent variable may have different determinants
- Implementation:
  - Generate the dummy variable $D = \begin{cases} 1 & \text{if the individual belongs to } G_A \\ 0 & \text{if the individual belongs to } G_B \end{cases}$
  - Estimate the original model 'duplicated':

$$Pr(Y|X) = F(\theta_0 + \theta_1 X_1 + \cdots + \theta_k X_k + \gamma_0 D + \gamma_1 D X_1 + \cdots + \gamma_k D X_k)$$

  - Apply a LR / Wald test for the significance of the variables where D is present:

    $H_0: \gamma_0 = \gamma_1 = \cdots = \gamma_k = 0$ (no structural break)

    H_1: Não H_0 (with a structural break)

Nonlinear models with panel data

Base nonlinear model for panel data:

- Individual effects model:

$$E(Y_{it}|x_{it}, \alpha_i) = G(\alpha_i + x_{it}'\beta)$$
$$Pr(Y_{it}|x_{it}, \alpha_i) = F(\alpha_i + x_{it}'\beta)$$

- Unlike the linear case:
  - Assuming $E(\alpha_i|x_{it}) = 0$ is not enough to get consistent estimators
  - In general, methods based on subtracting time averages or first-diferences do not eliminate fixed effects
  - Inconsistent estimation of α_i leads to inconsistent estimation of β (incidental parameters problem)

Nonlinear models with panel data

Main estimators:

- Pooled estimator:
  - Based on the estimation of the model $E(Y_{it}|x_{it}) = G(x'_{it}\beta)$ and $Pr(Y_{it}|x_{it}) = F(x'_{it}\beta)$: requires the assumption of no individual effects
  - Even with random effects this estimator will be, in general, inconsistent

Nonlinear models with panel data

- Fixed effects estimator:
  - Assumes $E(\alpha_i|x_{it}) \neq 0$
  - Long panels
    - Adds dummies for each individual, allowing estimation of the $\alpha_i's$
  - Short panels:
    - In a few cases:
      - It is possible to drop the $\alpha_i's$ from the model to be estimated using methods defined on a case-by-case basis (may also be used with long panels)
      - In general, prediction and quantification of partial effects are not possible
    - In most cases, no fixed effects estimator is available

Nonlinear models with panel data

- Random effects estimator:
  - Most popular panel data estimator for probabilistic models
  - It is necessary to:
    - Correctly specify $f(y_i|x_i, \alpha_i; \beta)$
    - Assume that α_i follows some distribution $f(\alpha_i; \eta)$
  - Density function for maximum likelihood estimation:

$$f(y_i|x_i; \beta, \eta) = \int f(y_i|x_i, \alpha_i; \beta) f(\alpha_i; \eta) d\alpha_i$$

    - In general, this expression cannot be simplified
    - Because of the integral, it requires numerical methods
  - QML estimation not available
  - In general, prediction and quantification of partial effects are not possible