# MICROECONOMETRICS II

2018/19

Esmeralda A. Ramalho

# Framework

## Course description:

- Addresses complementary topics relative to Microeconometrics
- 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

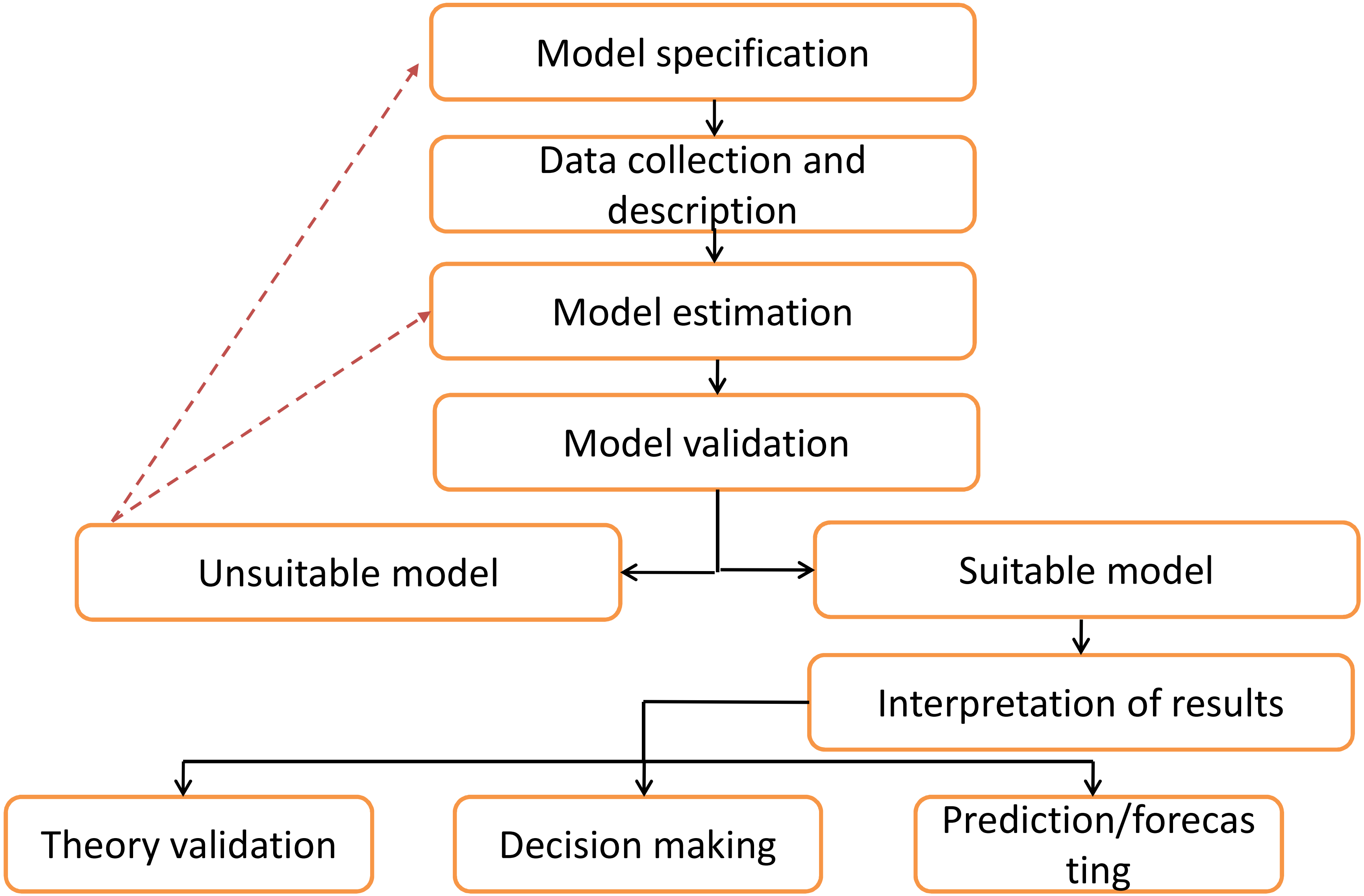
**Basic reference:** Cameron A.C. e Trivedi P.K. (2005), Microeconometrics, Methods and Applications, Cambridge

# Framework

## List of topics:

- Estimation and inference in nonlinear models: extremum estimators, GMM estimators, maximum likelihood estimators
- 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

# 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 numeric characteristics of the dependent variable restricts the variants that may be applied in each case:

| $Y$ | **Type of outcome** | **Main model** |
|---|---|---|
| $]-\infty,+\infty[$ | Unbounded data | Linear |
| $\{0,1\}$ | Binary choices | Logit,... |
| $\{0,1,2,\ldots,J-1\}$ | Multinomial choices | Multinomial logit,... |
| $\{0,1,2,\ldots,J-1\}$ | Ordered choices | Ordered logit,... |
| $\{0,1,2,\ldots\}$ | 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: $\Delta X_j = 1 \Longrightarrow$
  - $\Delta E(Y|X) = \frac{\partial E(Y|X)}{\partial X_j} = \frac{\partial G(X\beta)}{\partial X_j} = \beta_j \frac{\partial G(X\beta)}{\partial X\beta} = \beta_j g(x_i'\beta)$
  - $\Delta P(Y|X) = \frac{\partial P\ \ (Y|X)}{\partial X_j} = \frac{\partial F(X\beta)}{\partial X_j} = \beta_j \frac{\partial F(X\beta)}{\partial X\beta} = \beta_j f(x_i'\beta)$

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

# Model Transformations and Adaptations

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

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 $\theta$:

$$\theta = \underset{\theta \in \Theta}{\operatorname{argmax}} \, Q_n(\theta)$$

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

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

# Extremum estimators

Particular cases:

1. M Estimators: maximum likelihood like

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

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

Particular cases:

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

Particular cases:

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 $\theta_0$

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

2. Asymptotic normality - Theorem 3.4

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

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 $W = \text{plim}\widehat{W} = B^{-1}$ 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 4th 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 $\theta$ 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 $\theta$

# 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 : $plim\hat{\theta} = \theta_0$
- Assyptotic normality: $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N[0, A^{-1}BA'^{-1}]$ (since the Fisher information matrix equality does not hold (correspond to a GMM estimator with exact identification)
- Assyptotic efficiency requires the additional assumption of the correct specification of $V(Y|X)$. In these conditions $A = -A$, which implies the simplification of $A^{-1}BA'^{-1}$ 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 $\theta_r$ and $\theta_u$, respectively

Theorem 9.2: all the following test statistics converge in distribution to $\chi_r^2$:

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

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

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

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

# GMM hypothesis testing

Particular case: ML

$$W = nr(\hat{\theta}_u)'[\hat{R}\hat{C}^{-1}\hat{R}']r(\hat{\theta}_u)$$
$$LM = n\nabla_\theta LL(\hat{\theta}_r)'\hat{C}\nabla_\theta LL(\hat{\theta}_r)$$
$$LR = n[LL(\hat{\theta}_r) - LL(\hat{\theta}_u)]$$

where $\hat{C}$ is the estimator of the information matrix

# GMM hypothesis testing

## Particular case: ML

- Test for the joint significance of a set of parameters: LR test
  - Restricted model: $Pr(Y|X) = F\left(\theta_{r0} + \theta_{r1}x_1 + \cdots + \theta_{rg}x_g\right)$
  - Full model: $Pr(Y|X) = F\left(\theta_0 + \theta_1 x_1 + \cdots + \theta_g x_g + \theta_{g+1}x_{g+1} + \cdots + \theta_k x_k\right)$
  - Hypothesis in test:

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

    $H_1$: No $H_0$ (full model)

- Both the restricted and unrestricted models have to be estimated

# GMM hypothesis testing

Particular case: ML

- Test for individual significance: Wald test
  - In general

$$W = \hat{\theta}_F' \left[ \text{Var}(\hat{\theta}_r) \right]^{-1} \hat{\theta}_F \sim \chi^2_{k-g}$$

where $\hat{\theta}_F = \left[ \theta_{g+1}, \dots, \theta_k \right]$, but for $H_0$: $\theta_j = 0$, $W$ simplifies to:

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

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

# GMM hypothesis testing

## Particular case: ML

- Score/LM test:

$$\text{Score} = \nabla_\theta LL(X\theta)\Big|_{\theta=\hat{\theta}_R} [\text{Var}(\theta)]^{-1} \nabla_\theta LL(X\theta)\Big|_{\theta=\hat{\theta}_R} \sim \chi^2_{k-g}$$

  - 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

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

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

# 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 $\theta$ also $\theta_1$ 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.

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

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

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