Advanced Econometrics

University of Warsaw

Jerzy Mycielski

Doctoral School of Social Sciences, 2024

Interpretation of econometric models.

- Structural: structure of the model is valid, parameters are unknown, likelihood base analysis (Maximum Likelihood Estimation, Bayesian estimators)
- Semiparametric: part of the probabilistic model is left unspecified (OLS, Generalised Method of Moments (GMM), semiparametric estimators)
- Quasi-structural approach: model is treated as an approximation rather than truth (quasi-likelihood function, quasi-MLE, and quasi-likelihood)
- Calibration approach: treat the models as approximations, rejects statistical methods of estimation, uses *ad hoc* methods to fit models

Types of empirical data.

- Experimental: values of explanatory variables are determined by researcher and therefore can be treated as nonrandom
- Observational: neither explained nor explanatory variables are under control of researcher
- Most of the empirical data in economics is observational
- It is difficult to infer the causal mechanism from observational data as we cannot manipulate one variable to see the effect of this changes on other variables.
- What we are observing in practice are co-movements (correlations) of the observed variables
- However: co-movements (correlations) do not imply causality!

Standard data structures.

- Cross-section: observations independent, large number of observations
  - Example: Labor Force Survey (LFS)
- Time-series: serial dependence, sampling freqency often low (annual, quarterly), number of observations low
  - Example: GDP, interest rates
- Panel: for individual units, data is collected for several time periods, combine characteristics of the cross-section and time-series data
  - Example: Panel Study of Income Dynamics (PSID)
- Clustered: observation within clusters dependent but clusters independent
  - household data
- Spatial: observations dependent but nature of dependence known and related to distance
  - Example: regional data

The Probability approach in econometrics

- Deterministic models cannot precisely describe empirical data
- It is obvious that there are random factors influencing actions of economic agents
- On the other hand economic agents when making decisions are taking into account random factors
- Therefore models used in analysis of empirical data should explicitly include random elements
- Models used in the analysis of the data should be probabilistic

Distributions and unconditional expectations

- Cumulative distribution function

$$F(u) = \mathbb{P}[wage < u]$$

- Density function

$$f(u) = \frac{d}{du} F(u)$$

- Median

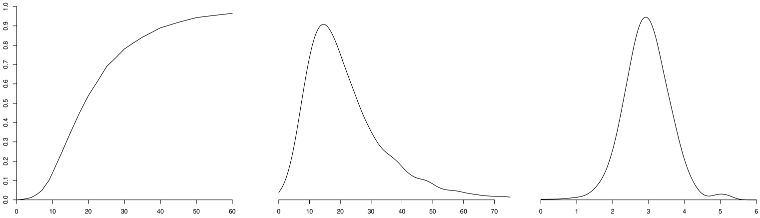
$$F(m) = \frac{1}{2}$$

- Expectation (mean)

$$\mathbb{E}[Y] = \sum_{j=1}^{\infty} \tau_j P(Y = \tau_j)$$

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f(y)\, dy$$

Distribution of wages



(a) Distribution (b) Density (c) Density of log(*wage*)

Figure 2.1: Wage Distribution and Density. All Full-time U.S. Workers

Source: Hansen (2022)

- Sample of 50,742 full-time non-military wage-earners reported in the March 2009 Current Population Survey
- Notice *log* (*wage*) has distribution much closer to normal distribution

Conditional expectation function (CEF)

- Expected values for subgroups

$$\mathbb{E}\left[log\left(wage \right) \middle| gender = man \right] = 3.05$$

$$\mathbb{E}\left[log\left(wage \right) \middle| gender = woman \right] = 2.81$$

- $log\left(wage \right)$ differential

$$\begin{aligned} & \mathbb{E}\left[log\left(wage \right) \middle| gender = man \right] \\ -& \mathbb{E}\left[log\left(wage \right) \middle| gender = woman \right] = 0.24 \end{aligned}$$

- $log\left(wage \right)$ differential can be interpreted percentage difference between wages of women and men.
- Indeed for $\frac{y_1}{y_2}$ close to 1:

$$log\left(y_1 \right) - log\left(y_2 \right) = log\left(\frac{y_1}{y_2} \right) \approx \frac{y_1}{y_2} - 1 = \frac{y_1 - y_2}{y_2}$$

Average wages by race and gender

Table 2.1: Mean Log Wages by Gender and Race

	men	women
white	3.07	2.82
Black	2.86	2.73
other	3.03	2.86

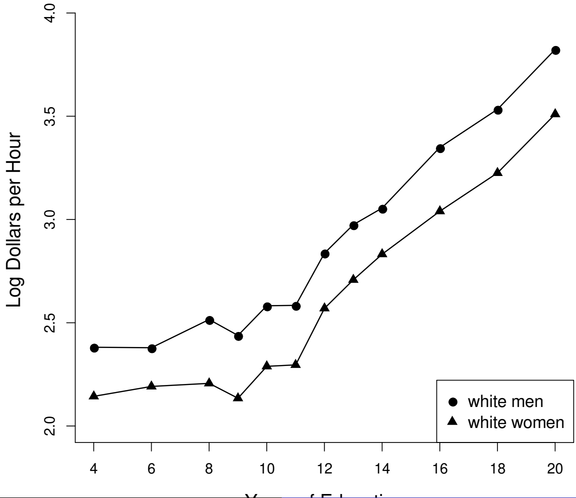
Source: Hansen (2022)

- Conditional means with given race and gender e.g.:

$$\mathbb{E}\left[log\left(wage\right) \middle| gender = man, race = Black \right] = 3.05$$

- We describe 50,742 with just 6 numbers

Conditional expectation function (CEF)



Conditional expectation function (CEF)

- Generic notation for conditional expectations

$$\mathbb{E}\left[\, Y \mid X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k\right] = m\left(x_1, x_2, \ldots, x_k\right)$$

or

$$\mathbb{E}\left[\, Y \mid X = x\right] = m\left(x\right)$$

where $X{=}(x_1, \ldots, x_k)'$ is a column vector of conditioning variables

- For given joint density $f\left(y, x\right)$ marginal density of x is

$$f_X\left(x\right) = \int_{-\infty}^{\infty} f\left(y, x\right) dx$$

- Conditional density for all x such that $f_X\left(x\right) > 0$ is defined as

$$f_{Y|X}\left(y \mid x\right) = \frac{f\left(y, x\right)}{f_X\left(x\right)}$$

Conditional expectation function (CEF)

- Conditional expectation

$$m(x) = \mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy$$

- Intuitively $\mathbb{E}[Y|X = x]$ is the expected value of Y for idealized sub-population in which $X = x$
- $m(x)$ is also known as regression function
- Variables included in vector X are called regressors
- The process of finding and interpreting $m(x)$ is known as regression analysis

Conditional expectation function (CEF)

- CEF is widely used in econometric analysis because it can be intepreted as prediction of Y given value of X
- Moreover it can be shown that in some sense CEF is the best prediction
- Some properties of predictors based on CEF are easy to derive
- It is not only used in econometrics but also in the economic theory: Rational Expactations Theory (RE)
- In the case when $m(X)$ is linear in X it is easy to estimate of $m(X)$

Properties of conditional expectations

Theorem

Simple Law of Iterated Expectations
If $\mathbb{E}\,|Y| < \infty$ *then for any random vector* X, $\mathbb{E}\,[\mathbb{E}\,[\,Y|\,X]] = \mathbb{E}\,[Y]$.

Theorem

Law of Iterated Expectations
If $\mathbb{E}\,|Y| < \infty$ *then for any random vectors* X_1 *and* X_2 ,
$\mathbb{E}\,[\mathbb{E}\,[\,Y|\,X_1, X_2]|\,X_1] = \mathbb{E}\,[\,Y|\,X_1]$.

Theorem

Conditioning Theorem
If $\mathbb{E}\,|Y| < \infty$ *then* $\mathbb{E}\,[g\,(X)\;Y|\,X] = g\,(X)\,\mathbb{E}\,[\,Y|\,X]$. *If in addition* $\mathbb{E}\,[g\,(X)] < \infty$ *then* $\mathbb{E}\,[g\,(X)\;Y] = \mathbb{E}\,[g\,(X)\,\mathbb{E}\,[\,Y|\,X]]$.

CEF error

- Define the difference between the observation for Y and prediction $m(X) = \mathbb{E}[Y|X]$ as prediction error e

$$e = Y - m(X)$$

- By definition

$$Y = m(X) + e$$

- Then

$$\begin{aligned} \mathbb{E}[e|X] &= \mathbb{E}[(Y - m(X))|X] \\ &= \mathbb{E}[(Y)|X] - \mathbb{E}[m(X)|X] \\ &= m(X) - m(X) = 0 \end{aligned}$$

- Using Simple Law of Iterated Expectations

$$\mathbb{E}[e] = \mathbb{E}[\mathbb{E}(e|X)] = \mathbb{E}(0) = 0$$

Regression variance

- Dispersion of e is related to precision of prediction of Y obtained from $m(x)$
- The standard measure of dispersion is variance

$$\sigma^2 = var[e] = \mathbb{E}\left[(e - \mathbb{E}[e])^2\right] = \mathbb{E}\left[e^2\right]$$

- f $\mathbb{E}\left[Y^2\right] < \infty$ then $\sigma^2 < \infty$.
- If $\mathbb{E}\left[Y^2\right] < \infty$ then

$$var[Y] \geq var[Y - E[Y|X1]] \geq var[Y - E[Y|X1, X2]].$$

- Adding the regressors always reduces regression variance!

Best predictor

- One of the aims of the regression analysis is explain (predict) the changes of Y with the changes X
- Is the conditional expectation a good way to make predictions about Y?

Theorem

Conditional Expectation as Best Predictor
If $E\,\mathbb{E}\left[Y^2\right] < \infty$ *, then for any predictor* $g\left(X\right)$*,*

$$\mathbb{E}\left[\left(Y - g\left(X\right)\right)^2\right] \geq \mathbb{E}\left[\left(Y - m\left(X\right)\right)^2\right]$$

where $m\left(X\right) = \mathbb{E}\left[\,Y|\,X\right]$*.*

- Regression function $m\left(X\right)$ produces the best predictions of Y!

Conditional variance and conditional standard error

- If $\mathbb{E}\left[e^2\right] < \infty$, the conditional variance of the regression error e
$$\sigma^2(x) = var\left[e|\, X = x\right] = \mathbb{E}\left[e^2 \middle| X = x\right]$$
.
- The conditional variance of e treated as a random variable is $var\left[e|\, X\right] = \sigma^2(X)$.
- Conditional standard error $\sigma^2(x) = \sqrt{\sigma^2(x)}$
- Variance decomposition: if $\mathbb{E}\left[X^2\right] < \infty$ then
$$var\left[X\right] = \mathbb{E}\left[var\left[X|\, W\right]\right] + var\left[E\left[X|\, W\right]\right].$$
- As $E\left[e|\, X\right] = 0$ then $var\left[e\right] = \mathbb{E}\left[var\left[e|\, X\right]\right]$
- The error is homoskedastic if $\sigma^2(x) = \sigma^2$ does not depend on x
- Homoscedasticity is an exception rather than a rule - in most empirical models error is heteroskedastic

Regression derivative (marginal effect) and ceteris paribus assumption

- Define the following measure of the reaction of the $m(x)$ to changes in x_1:

$$\nabla_1 m(x) = \begin{cases} \frac{\partial m(x_1,\ldots,x_k)}{\partial x_1}, & \text{if } X_1 \text{ is continuous} \\ m(1,\ldots,x_k) - m(0,x_2,\ldots,x_k), & \text{if } X_1 \text{ is binary} \end{cases}$$

- Using the same definition for all elements of x we can define

$$\nabla m(x) = \begin{bmatrix} \nabla_1 m(x) \\ \vdots \\ \nabla_k m(x) \end{bmatrix}$$

- Notice that $\nabla_1 m(x)$ is defined as partial derivative of $m(x)$, therefore it measures the change of $m(x)$ if x_1 changes but all other variables are kept constant (ceteris paribus)
- This measure make possible to disentangle the influence of x_1 from the influence of other regressors (colliders and mediators)!

Linear CEF model

- Assume that CEF is the following function

$$m(x) = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_{k-1} + \beta_k = x'\beta.$$

- β_k is known as constant term or intercept
- $x = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$

$$\begin{aligned} Y =& X'\beta + e \\ E\left(e|\, X\right) =& 0 \end{aligned}$$

- Additionaly homoscedasticity assumption $var\left(e|\, X\right) = \sigma^2$ is sametimes added

Linear CEF with nonlinear effects, with dummy variables

- Quadratic regression

$$m(x1, x2) = x_1\beta_2 + x_2\beta_2 + x_1^2\beta_3 + x_2^2\beta_4 + x_1x_2\beta_5 + \beta_6$$

- It is nonlinear in the regressors x_1, x_2 but it is linear in the coefficients.
- Regression derivative w.r.t. x_1

$$\frac{\partial m(x1, x2)}{\partial x_1} = \beta_2 + 2x_1\beta_3 + x_2\beta_5$$

- β_5 is related to so called interaction effect
- Qualitative variables are coded using binary (dummy) variables

$$\begin{cases} X_1 = 1 & \text{if } gender = man \\ X_1 = 0 & \text{if } gender = woman. \end{cases}$$

- If qualitative variable has more than $s > 2$ possible values we use $s - 1$ dummy variables to represent it

Best linear predictor

- Assume $\mathbb{E}\left[Y^2\right] < \infty$, $\mathbb{E}\left\|X\right\|^2 < \infty$, $\mathbb{E}\left[XX'\right]$ is positive semidefinite (invertible)
- Linear Predictor of Y given X is $\mathscr{P}\left[Y|X\right] = X'\beta$ where β minimizes the mean squared prediction error:

$$S(\beta) = \mathbb{E}\left[\left(Y - X'\beta\right)^2\right].$$

- Solving this minimization problem gives the following formula for β (Linear Projection Coefficient)

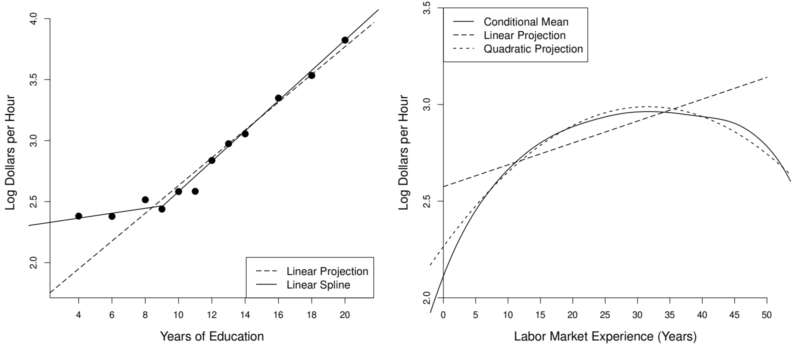
$$\beta = \left(\mathbb{E}\left[XX'\right]\right)^{-1}\mathbb{E}\left[XY\right]$$

- Then the Best Linear Predictor (Linear Projection) is given by

$$\mathscr{P}\left[Y|X\right] = X'\left(\mathbb{E}\left[XX'\right]\right)^{-1}\mathbb{E}\left[XY\right]$$

- In the linear projection model $Y = X'\beta + \alpha + e$, $\mu_Y = \mathbb{E}(Y)$, $\mu_X = \mathbb{E}(x)$

$$\begin{aligned} \alpha =& \mu_Y - \mu_X\beta \\ \beta =& var\left[X\right]^{-1} cov\left(X, Y\right). \end{aligned}$$



(a) Projections onto *education*

(b) Projections onto *experience*

Figure 2.6: Projections of log(*wage*) onto *education* and *experience*

Source: Hansen (2022)

Omitted variable bias

- Consider the following regression model

$$Y = X_1'\beta_1 + X_2'\beta_2 + e$$

- Make Linear Projection of $\dot{Y}$on X_1 only
- In such a case Linear Projection Coefficient is

$$\begin{aligned}
\gamma_1 =& \mathbb{E}\left[X_1 X_1'\right]^{-1} \mathbb{E}\left[X_1 Y\right] \\
=& \mathbb{E}\left[X_1 X_1'\right]^{-1} \mathbb{E}\left[X_1 \left(X_1'\beta_1 + X_2'\beta_2 + e\right)\right] \\
=& \beta_1 + \mathbb{E}\left[X_1 X_1'\right]^{-1} \mathbb{E}\left[X_1 X_2'\right] \beta_2 = \beta_1 + \Gamma_{12}\beta_2
\end{aligned}$$

- Generally speaking $\gamma_1 = \beta_1 + \Gamma_{12}\beta_2 \neq \beta_1$, the coefficient is biased estimate of β_1!

Misspecified model

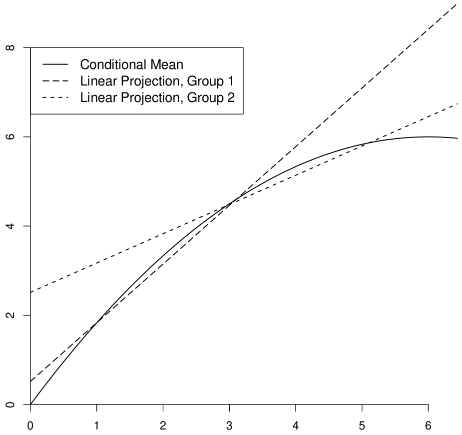


Figure 2.7: Conditional Mean and Two Linear Projections

Source: Hansen (2022)

CEF and Causal effect

- Model

$$Y = h(D, X, U)$$

- X are observable factors, U unobservable factors
- Causal effect of D on Y is

$$C(X, U) = Y(1) - Y(0) = h(1, X, U) - h(0, X, U)$$

interpreted as change in Y due to treatment while holding U constant

- The conditional average causal effect o of D on Y is

$$ACE(x) = \mathbb{E}[C(X, U)| X = x] = \int_{\mathbb{R}^l} C(x, u) f(u| x) du$$

where $f(u)$ is the density of U.

- The unconditional average causal effect of D on Y is

$$ACE = \mathbb{E}[C(X, U)] = \int ACE(x) f(x) dx$$

Conditional Independence Assumption (CIA)

- We say that variables U and D are conditionally independent if conditional on X the random variables D and U are statistically independent

$$f\left(u|\, D, X\right) = f(u|X)$$

- In such a case

$$\begin{aligned} m\left(d,x\right) =& \mathbb{E}\left[\left.Y\right| D=d, X=x\right] = \mathbb{E}\left[\left.h\left(d,x,U\right)\right| D=d, X=x\right] \\ =& \int h\left(d,x,u\right) f\left(\left.u\right| d,x\right) du. = \int h\left(d,x,u\right) f\left(\left.u\right| x\right) du. \end{aligned}$$

- Therefore

$$\begin{aligned} \nabla m\left(d,x\right) =& m\left(1,x\right) - m\left(0,x\right) \\ =& \int h\left(1,x,u\right) f\left(\left.u\right| x\right) du - \int h\left(0,x,u\right) f\left(\left.u\right| x\right) du \\ =& C\left(x,u\right) f\left(\left.u\right| x\right) du = ACE\left(x\right) \end{aligned}$$

- CIA implies $\nabla m\left(d,x\right) = ACE(x)$!

Moment estimators

- Sample is the set $\{(Y_i, X_i) : i = 1, ..., n\}$ of n realisations of the random variables (Y, X)
- We often assume that elements of the sample are identically distributed that is they are draws from common distribution F
- In econometric theory we refer to the underlying common distribution F as the population or Data Generating Prosess (DGP)
- The simplest estimators are based on moments that is by replacing population moments by sample moments
- E.g. expected value $\mu =$ and variance σ^2 of Y can be estimated as follows:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} Y_i}{n}, \ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right]^2 = \frac{\sum_{i=1}^{n} (Y_i - \hat{\mu})^2}{n}$$

as $var(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2$

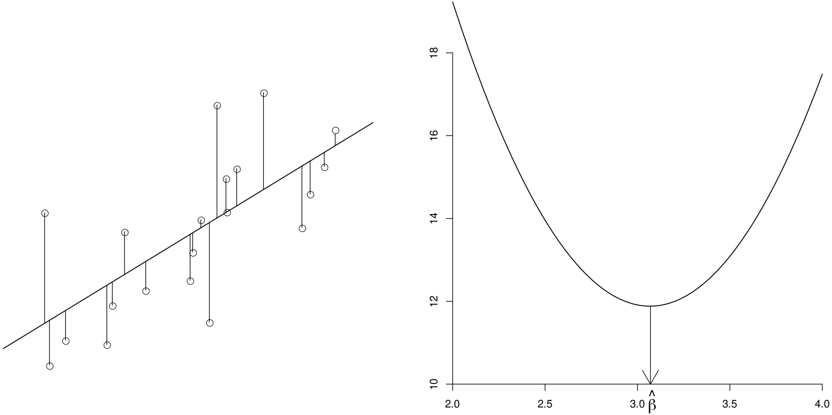
Linear CEF estimation

- Replacing the variance of CEF error with sample variance:

$$\hat{S}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - X_i'\beta\right)^2 = \frac{1}{n} SSE\left(\beta\right)$$

  where $SSE\left(\beta\right)$ is called the sum of squared errors function.
- We define the least squares estimator $\hat{\beta}$ as the minimizer of $\hat{S}(\beta)$

Sum of Squared Error, one regressor



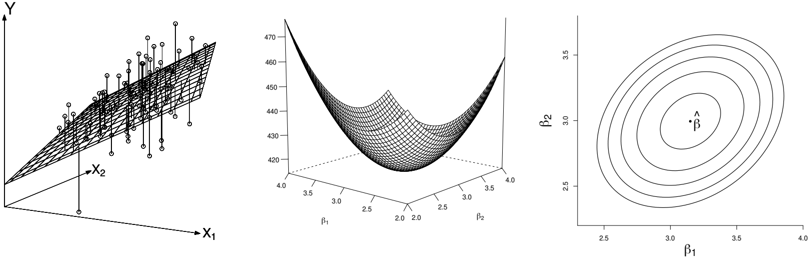
(a) Deviation from Fitted Line

(b) Sum of Squared Error Function

Figure 3.1: Regression With One Regressor

Source: Hansen (2022)

Sum of Squared Error, two regressors



(a) Regression Plane

(b) Sum of Squared Error Function

(c) SSE Contour

Figure 3.2: Regression with Two Variables

Source: Hansen (2022)

Bibliography

Hansen, B. (2022). *Econometrics*. Princeton University Press. ISBN: 9780691235899.