Regression review for UrbanSIM control function project

Danqing Zhang

March 1, 2016

Contents

1 OLS estimator in econometrics [1]

1.1 assumptions of OLS estimator

- Zero mean

$$E(\epsilon) = 0 \tag{1}$$

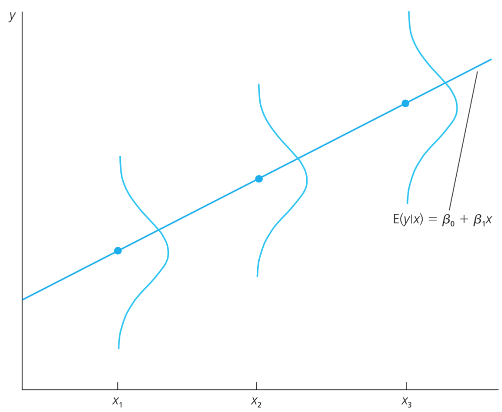
- Zero conditional mean

$$E(\epsilon|x) = 0 \tag{2}$$

$$E(y|x) = \theta^\top x + b \tag{3}$$

FIGURE 2.1

E($y|x$) as a linear function of x.



The above assumptions are used to motivate OLS estimators of given a random sample of data.

1.2 Derive OLS estimator from assumptions

There are several ways to motivate the following estimation procedure. We will use zero mean assumption and an important implication of zero conditional mean assumption:

with x. Therefore, we see that u has zero expected value and that the *covariance* between x and u is zero:

$$\mathrm{E}(u) = 0 \tag{2.10}$$

and

$$\mathrm{Cov}(x,u) = \mathrm{E}(xu) = 0, \tag{2.11}$$

Then change the above equations in terms of the observable variables x and y and the unknown weight vector θ:

$$\mathrm{E}(y - \beta_0 - \beta_1 x) = 0 \tag{2.12}$$

and

$$\mathrm{E}[x(y - \beta_0 - \beta_1 x)] = 0, \tag{2.13}$$

β_0 and β_1. In fact, they can be. Given a sample of data, we choose estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ to solve the *sample* counterparts of (2.12) and (2.13):

$$n^{-1}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{2.14}$$

and

$$n^{-1}\sum_{i=1}^{n}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0. \tag{2.15}$$

In multivariate case, the equation is expressed as below:

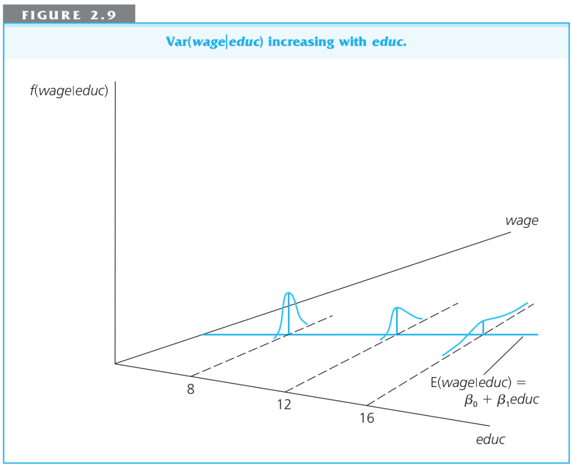
$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{\theta}^\top x_i) = 0 \tag{4}$$

$$\frac{1}{N}\sum_{i=1}^{N}x_{ik}(y_i - \bar{\theta}^\top x_i) = 0, \forall k \tag{5}$$

which is exactly the first order conditions of the unconstrained optimization problem:$\min_{\theta} \sum_{i=1}^{N}(y^i - \theta^T x^i)^2$. So that means, if we can select input variables that can fit the zero mean assumptions and zero conditional mean assumptions, then it naturally meet the OLS requirements, so we can get unbiased estimator.

1.3 statistical properties of OLS estimator

Homoskedasticity:
that is:
$var(\epsilon|x) = \sigma^2$ for univariate case
And $var(\epsilon|x) = \Sigma$ for multivariate case
We must emphasize that the homoskedasticity assumption is quite distinct from the zero conditional mean assumption
Why we need the Homoskedasticity assumption?:
it is important to know how far we can expect $\bar{\theta}$ to be away from *theta* on average. Among other things, this allows us to choose the best estimator among all, or at least a broad class of, unbiased estimators.
case where the Homoskedasticity is not met:



2 MLE estimator in ML

2.1 Notations & Normal Distribution

n:dimensions of x^i
N:size of data

$$x = \begin{bmatrix} x_0 \\ x_1 \\ . \\ . \\ x_n \end{bmatrix} \in R^{n+1} \ \Theta = \begin{bmatrix} \Theta_0 \\ \Theta_1 \\ . \\ . \\ \Theta_n \end{bmatrix} \in R^{n+1}$$

$y = \theta^T x + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \Sigma)$
Then:
$y^i = \theta^T x^i + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \Sigma)$
Then:
$P(y^i|\theta, x^i, \Sigma) = N(\theta^T x^i, \Sigma)$
Then:

$P(y|\theta, X, \Sigma) = \prod_{i=1}^{N} P(y^i|\theta, x^i, \Sigma)$

If $y^i \sim N(\mu^i, \Sigma)$,then:
$P(y^i|\mu^i, \Sigma) = |2\pi\Sigma|^{\frac{1}{2}} exp(-\frac{1}{2}(y^i - \mu^i)^T \Sigma^{-1}(y^i - \mu^i))$

2.2 Why we do not consider the assumptions problem in ML?

In ML we make the strongest assumption that,$\epsilon \sim \mathcal{N}(0, \Sigma)$, and we make the strongest assumption that x and u are independent

Then given the above two strongest assumptions, the three assumptions zero mean, zero conditional mean,Homoskedasticity will naturally met.

2.3 MLE = OLS in ML framework assumptions

Typically we can use MLE and least sqaure method to compute μ, both of which in the end are optimization problems
We here need to prove that the two method are in essence the same:
$Ln(P(y|\theta, X, \sigma^2)) = \sum_{i=1} N Ln(P(y^i|\theta, x^i, \sigma^2)) = \sum_{i=1}^{N} Ln(N(y^i|\theta, x^i, \sigma^2))$
$= \sum_{i=1}^{N}[-\frac{1}{2}Ln(2\pi\sigma^2)] - [\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y^i - \theta^T x^i)^2]$

Then $\underset{\theta}{min} L(\theta)$ is identical to $\underset{\theta}{min} \frac{1}{2}\sum_{i=1}^{N}(y^i - \theta^T x^i)^2$

2.4 Difference

Economics deal with variable selection, and interpreting the weights,and naturally how to deal with the situations where the three assumptions do not hold(also random sampling is another assumption that should be met, when it is not satisfied, certain methods shall be used).There we need intuition of economic phenomena as well as some interpretation of data analysis result of different regression models

And machine learning [4] is after the variable selection step, when we have successfully select variables of interest. We improve the models by :(1) just regression-related model:we seek to find the best estimator for prediction, so to prevent over-fitting, we use regularization method;(2) when predicting two variables(input,output) that are macro-variables of an integrated system, it is better to use state-space model, then the framework of MLE framework(or MAP in Bayesian)to convert the parameter estimation(identification) problem into an optimization problem is still applicable. And that is why ML is so powerful.

3 Common problems and methods in Economics when the four assumptions are not satisfies(zero mean, zero conditional mean, Homoskedasticity, random sampling)

3.1 Homoskedasticity is not satisfied [2]

heteroscedasticity problem

3.2 Random sampling assumption is not satisfied [2]

3.3 Zero condition mean assumption is satisfied:endogenous problem [2]

Types of reason of violation of zero condition mean assumptions:

- Functional Form Misspecification
- Models with Random Slopes
- Missing variables
- reverse causality

Possible solutions:

- Proxy Variables for Missing variables
- instrumental variable for all the other cases

3.4 Week Exdogenity [3]

1. **weak exogeneity:** Let $p(y_i, \mathbf{x}_i)$ be the joint probability density of y_i and $\mathbf{x}_i$, where $i = 1, \ldots, n$ indexes the observations. That is, both $\mathbf{y}$ and $\mathbf{X}$ are considered random variables. Since a joint density can always be factored as the product of a conditional density and a marginal density, we have

$$p(y_i, \mathbf{x}_i|\boldsymbol{\theta}) = p(y_i|\mathbf{x}_i, \boldsymbol{\theta})\,p(\mathbf{x}_i|\boldsymbol{\theta}).$$

100 GETTING STARTED: BAYESIAN ANALYSIS FOR SIMPLE MODELS

The assumption of weak exogeneity consists of the following two restrictions:

$$p(y_i|\mathbf{x}_i, \boldsymbol{\theta}) = p(y_i|\mathbf{x}_i, \boldsymbol{\theta}_{y|x})$$

and

$$p(\mathbf{x}_i|\boldsymbol{\theta}) = p(\mathbf{x}_i|\boldsymbol{\theta}_x),$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_{y|x}, \boldsymbol{\theta}_x)'$. That is, without loss of generality, the parameter vector $\boldsymbol{\theta}$ can be decomposed into two components, one indexing the conditional density $p(y_i|\mathbf{x}_i)$ and the other indexing the marginal density $p(\mathbf{x}_i)$. An implication of these assumptions (and this is the substantive content of weak exogeneity) is that knowledge of the parameters $\boldsymbol{\theta}_x$ indexing $p(\mathbf{x}_i)$ provides no additional information

4 Regression with Time Series Data [1]

Part 2: chapter 10-12

5 Formulation of logistics regression

5.1 Theorem of EV

`Generalized Extreme Value Distribution`

if $\epsilon \sim EV(\eta, \mu)$, then $a\epsilon + b \sim EV(a\eta + b, \frac{\mu}{a})$, from here we can get the theorem for two competing EV variable, just like two exponentially- distributed variables.

$$Ifx_1 \sim exp(\lambda_1), x_2 \sim exp(\lambda_2), then : P(x_1 < x_2) = \frac{\lambda_1}{\lambda 1 + \lambda_2} \tag{6}$$

5. If $\varepsilon_a \sim \mathrm{EV}(\eta_a, \mu)$ and $\varepsilon_b \sim \mathrm{EV}(\eta_b, \mu)$ are independent with the same scale parameter μ, then

$$\varepsilon = \varepsilon_a - \varepsilon_b \sim \mathrm{Logistic}(\eta_a - \eta_b, \mu),$$

namely

$$f(\varepsilon) = \frac{\mu e^{-\mu(\varepsilon - \eta_a + \eta_b)}}{(1 + e^{-\mu(\varepsilon - \eta_a + \eta_b)})^2}, \tag{4.25}$$

$$F(\varepsilon) = \frac{1}{1 + e^{-\mu(\varepsilon - \eta_a + \eta_b)}}, \quad \mu > 0, -\infty < \varepsilon < \infty. \tag{4.26}$$

$$\tag{4.27}$$

6. If $\varepsilon_i \sim \mathrm{EV}(\eta_i, \mu)$, for $i = 1, \ldots, J$, and ε_i are independent with the same scale parameter μ, then

$$\varepsilon = \max_{i=1,\ldots,J} \varepsilon_i \sim \mathrm{EV}(\eta, \mu) \tag{4.28}$$

where

$$\eta = \frac{1}{\mu} \ln \sum_{i=1}^{J} e^{\mu \eta_i}. \tag{4.29}$$

It is important to note that this property holds only if all ε_i have the same scale parameter μ. As ε follows an extreme value distribution, its expected value is

$$E[\varepsilon] = \eta + \frac{\gamma}{\mu}.$$

Equivalently,

$$\eta = E[\varepsilon] - \frac{\gamma}{\mu}.$$

5.2 from assumption of $\epsilon \sim EV(\eta, \mu)$ to expression of $P(t_i = 1|\theta, b)$, and then to MLE estimation of parameters

$$U_1 = \theta^\top x + b + \epsilon, \epsilon \sim EV(\eta, \mu) \tag{7}$$

$$U_0 = \epsilon, \epsilon \sim EV(\eta, \mu)$$

Then how to get

$$P(t = 1) = \frac{\exp(\theta^\top x + b)}{1 + \exp(\theta^\top x + b)} \tag{8}$$

from above definition of probabilistic utility function and theorem of EV(extreme values)?

References

[1] Jeffrey Wooldridge. *Introductory econometrics: A modern approach.* Nelson Education, 2015.

[2] Dandan Zhang. Peking university national school of development intermediate econometrics, 2013.

[3] Simon Jackman. *Bayesian analysis for the social sciences*, volume 846. John Wiley & Sons, 2009.

[4] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.