

## Week 5: Estimators and Estimation Methods, Nonlinear Regression, Quantile Regression

Advanced Econometrics 4EK608

Vysoká škola ekonomická v Praze

# Outline

- 1 Estimators and estimation methods
  - Properties of estimators - repetition from BSc courses
  - Method of moments
  - Maximum likelihood estimation
- 2 Nonlinear regression models
- 3 Quantile regression

# Estimators and estimation methods

Notation:

- $\theta$  - population parameter
- $(x_1, x_2, \dots, x_n)$  - random sample of  $n$  observation of  $x$
- $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$  is an estimator of  $\theta$

Basic notions:

- All estimators posses sampling distribution  
mean:  $\mathbf{E}(\hat{\theta})$   
variance:  $\mathbf{E}[(\hat{\theta} - \mathbf{E}(\hat{\theta}))^2]$   
etc.
- Estimators  $\times$  estimate
- Many estimators exist for a parameter (population mean):

$$\hat{\theta}_1 = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\hat{\theta}_2 = \tilde{x} = \frac{1}{2}(x_{max} + x_{min})$$

# Estimators and estimation methods

Small sample properties of estimators & definitions:

- **Unbiasedness:** the mean of sampling distribution equals the parameter being estimated
- **Efficiency:** an estimator is efficient if it is unbiased and no other unbiased estimator has a smaller variance. This is usually difficult to prove, that is why we simplify the concept:
  - Relative efficiency
  - Linear unbiased estimators instead of unbiased estimators (linear estimator is linear function of sample observations)

# Estimators and estimation methods

Small sample properties of estimators & definitions:

Best Linear Unbiased Estimator (BLUE) is linear, unbiased and no other linear unbiased estimator has a smaller variance. It is not necessarily the best estimator.

- Non-linear estimators can be better
- Biased estimators can have smaller Mean Square Error: sum of variance and the squared bias

# Estimators and estimation methods

Large sample properties of estimators & definitions:

- Sampling distribution of an estimator changes with the size of sample.
- Asymptotic distribution for any estimator is that distribution to which the sampling distribution tends as the sample becomes larger. Its  $1^{st}$  and  $2^{nd}$  moments are asymptotic mean and asymptotic variance.
- When the sampling distribution collapses onto a single value when the sample becomes larger, we call this value probability limit. We say estimator converges in probability to that value

# Estimators and estimation methods

Large sample properties of estimators & definitions:

- Asymptotic unbiasedness
- **Consistency**
- Unbiased estimators are not necessarily consistent.
- If  $\hat{\theta}$  is an unbiased estimator of  $\theta$  and  $\text{plim}(\text{var}(\hat{\theta})) = 0$  i.e.  $[\text{var}(\hat{\theta}) \rightarrow 0 \text{ as } n \rightarrow \infty]$ , then  $\text{plim}(\hat{\theta}) = \theta$ .
- Consistent estimators: unbiased & their variance shrinks to zero as sample size grows (entire population is used).
  - Minimal requirement for estimator used in statistics or econometrics.
  - If some estimator is not consistent, then it does not provide estimates of population  $\theta$  values, even with unlimited data.

# Estimators and estimation methods

Large sample properties of estimators & definitions:

- Asymptotic efficiency: An estimator is asymptotically efficient if it is asymptotically unbiased and no other asymptotically unbiased estimator has smaller asymptotic variance.
- Asymptotic efficiency is usually difficult to prove, that is why we simplify the concept:
  - Relative asymptotic efficiency
  - Linear asymptotically unbiased estimators instead of asymptotically unbiased estimators



# Estimators and estimation methods

## Method of moments

- With the method of moments, we simply estimate population moments by corresponding sample moments.
- Under very general conditions, sample moments are consistent estimators of the corresponding population moments, but NOT necessarily unbiased estimators.

### Application example 1

Sample covariance is a consistent estimator of population covariance.

### Application example 2

OLS estimators we have used for parameters in the CLRM can be derived by the method of moments.

# Estimators and estimation methods

## Method of moments (MM)

Population moments, stochastic variable  $X$

- $\mathbf{E}(X^r)$ :  $r^{th}$  population moment about zero
- $\mathbf{E}(X)$ : the population mean is the first moment about zero
- $\mathbf{E}[(X - \mathbf{E}(X))^2]$ : the population variance is the second moment about the mean

Sample moments, sample observations  $(x_1, x_2, \dots, x_n)$

- $\frac{\sum_{i=1}^n x_i^r}{n}$ :  $r^{th}$  sample moment about zero
- $\frac{\sum_{i=1}^n x_i}{n} = \bar{x}$ : sample mean is the first moment about zero
- $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ : sample variance is the second sample moment about the mean

# Estimators and estimation methods

- In a LRM:  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$ , the  $k + 1$  parameters are **OLS**-estimated by minimizing:

$$\sum_{i=1}^n \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik} \right)^2 \quad (1)$$

- 1<sup>st</sup> order conditions for (1), plus assumptions  $E(u) = 0$  and  $E(x_j \cdot u) = 0$ , can be combined into a **MM** estimator:

$$\sum_{i=1}^n \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik} \right) = 0$$

$$\sum_{i=1}^n x_{i1} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik} \right) = 0$$

...

$$\sum_{i=1}^n x_{ik} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik} \right) = 0$$

# Estimators and estimation methods

- Let  $\mathbf{h}(\mathbf{w}) = \mathbf{h}(y, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta})$  define a regression model, where  $\mathbf{z}$  is a set of  $r$  instruments (IVs) - see Week 8. For simplicity, we can start by assuming  $\mathbf{x} \equiv \mathbf{z}$ .
- Method of moments estimator**  $\hat{\boldsymbol{\theta}}_{MM}$  minimizes:

$$\min_{\hat{\boldsymbol{\theta}}} : \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{h}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}) \right]' \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{h}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}) \right]$$

- If  $\mathbf{x} \neq \mathbf{z}$  and  $\#$  IVs  $>$   $\#$  regressors (overidentification), **Generalized method of moments** is used (GMM)

$$\min_{\hat{\boldsymbol{\theta}}} : \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{h}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}) \right]' \mathbf{W}_n \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{h}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}) \right]$$

where  $\mathbf{W}_n$  is a conveniently chosen  $(r \times r)$  matrix. (any positive definite matrix that may depend on data but not on  $\theta$ , e.g.  $I_r$ . Optimum  $\mathbf{W}_n$ : see e.g. Greene, chapter. 13.4.2)

# Estimators and estimation methods

## MM - consistency conditions

- **Convergence of the moments:** Sample moments converge in probability to their population counterparts.
- **Identification:** Parameters are identified in terms of the moment equations.
  - **Order condition:**  $\# \text{ IVs} \geq \# \text{ model variables}$ .
  - **Rank condition:** Moment equations are not redundant
  - Identification will be discussed in detail during Week 8
- **Limiting Normal distribution for the sample moments:** Population moments obey central limit theorem (CLT) or some similar variant.

# Estimators and estimation methods

## MM - summary

- MM is robust to differences in “specification” of the data generating process (DGP). → i.e. sample mean or sample variance estimate their population counterparts (assuming they exist) regardless of DGP.
- MM is free from distributional assumptions.
- “Cost” of this approach: if we know the specific distribution of a DGP, MM does not make use of such information → inefficient estimates.
- Alternative approach: method of maximum likelihood utilizes distributional information and is more efficient (provided this information is valid).

# Estimators and Estimation Methods

## Maximum likelihood estimator

Single  $\theta$  parameter case:

- 1<sup>st</sup> step: deriving a likelihood function  
 $L = L(\theta, y_1, y_2, \dots, y_n)$ , where  $y_i$  is observation of  $Y$  (stochastic),  $\theta$  is parameter of the distribution.
- 2<sup>nd</sup> step: finding maximum of  $L$  with respect to  $\theta$ ,  
that maximum is  $\tilde{\theta} = \theta_{MLE}$

With more parameters:  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$

$$L = L(\theta_1, \theta_2, \dots, \theta_m, y_1, y_2, \dots, y_n)$$

We find MLEs of the  $m$  parameters by partially differentiating the likelihood function  $L$  with respect to each  $\theta$  and then setting all the partial derivatives obtained to zero.

# Estimators and estimation methods

Likelihood function:

$$f(y_1, y_2, \dots, y_n | \boldsymbol{\theta}) = \prod_{i=1}^n f(y_i | \boldsymbol{\theta}) = L(\boldsymbol{\theta} | \mathbf{y})$$

where  $f(y | \boldsymbol{\theta})$  is the pdf of  $y$ , conditioned on set of parameters  $\boldsymbol{\theta}$ .

Maximum likelihood estimation of CLRM parameters:

$$\begin{aligned} \text{CLRM: } y_i &= \alpha + \beta x_i + \varepsilon_i & \mathbf{E}(y_i) &= \alpha + \beta x_i = \mathbf{x}_i' \boldsymbol{\beta} \\ & & \text{var}(y_i) &= \text{var}(\varepsilon_i) = \sigma^2 \end{aligned}$$

Probability density function for **Normal distribution**:

$$f(y | \boldsymbol{\theta}) = (2\pi\sigma^2)^{-0.5} \exp[-(y - \mu)^2 / 2\sigma^2]$$

In the case of CLRM, for each  $y_i = \alpha + \beta x_i + \varepsilon_i$ :

$$f(y_i, \mathbf{x}_i | \boldsymbol{\theta}) = (2\pi\sigma^2)^{-0.5} \exp[-(y_i - \mathbf{E}(y_i))^2 / 2\sigma^2], \text{ that is:}$$

$$f(y_i, \mathbf{x}_i | \boldsymbol{\theta}) = (2\pi\sigma^2)^{-0.5} \exp[-(y_i - \mathbf{x}_i' \boldsymbol{\beta})^2 / 2\sigma^2]$$



# Estimators and estimation methods

Log-likelihood ( $LL$ ) function, **Normal distribution** assumed, estimation of CLRM parameters:

$$\begin{aligned}\log L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}) &= \sum_{i=1}^n \log[f(y_i|\mathbf{x}_i, \boldsymbol{\theta})] = \\ &= -\frac{1}{2} \sum_{i=1}^n \left\{ \log(2\pi) + \log(\sigma^2) + \frac{1}{\sigma^2} [y_i - \mathbf{x}_i' \boldsymbol{\beta}]^2 \right\} = \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - \mathbf{x}_i' \boldsymbol{\beta}]^2\end{aligned}$$

numerical iterative method is used for  $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2)$  estimation (by maximizing the log-likelihood function).

if CLRM assumptions hold  $\Rightarrow$  MLE estimators  $\tilde{\alpha}$ ,  $\tilde{\beta}$  and  $\tilde{\sigma}^2$  are identical to OLS-generated estimators  $\hat{\alpha}$ ,  $\hat{\beta}$  (and  $\hat{\sigma}^2$ ).

# Estimators and Estimation Methods

## Basic MLE assumptions

- **Parameter space:** Gaps and nonconvexities in parameter spaces would generally collide with estimation algorithms (settings such as  $\sigma^2 > 0$  are OK).
- **Identifiability:** The parameter vector  $\theta$  is identified (estimable), if for two vectors,  $\theta^* \neq \theta$  and for some data observations  $\mathbf{x}$ ,  $L(\theta^*|\mathbf{x}) \neq L(\theta|\mathbf{x})$ .
- **Well-behaved data:** Laws of large numbers (LLN) apply. Some form of CLT can be applied to the gradient (i.e. for the estimation method).
- **Regularity conditions:** “well behaved” derivatives of  $f(y_i|\theta)$  with respect to  $\theta$  (see Greene, chapter 14.4.1).

# Estimators and Estimation Methods

## MLE properties

- **Consistency:**  $\text{plim}(\hat{\theta}) = \theta_0$  ( $\theta_0$  is the true parameter)
- **Asymptotic normality** of  $\hat{\theta}$
- **Asymptotic efficiency:**  $\hat{\theta}$  is asymptotically efficient and achieves the Cramér-Rao lower bound for consistent estimators (see Greene, chapter 14.4.5)
- **Invariance:** MLE of  $\gamma_0 = c(\theta_0)$  is  $c(\hat{\theta})$  if  $c(\theta_0)$  is a continuous and countinuously differentiable function.

# Estimators and estimation methods

## MLE - summary

- MLE is only possible if we know the form of the probability distribution function for the population (Normal, Poisson, Negative Binomial, etc.).
- MLEs possess the large sample properties of consistency and asymptotic efficiency. There is no guarantee that they possess any desirable small-sample properties.
- Under CLRM assumptions, MLE estimator are identical to OLS estimators.
- MLE-related tests (Likelihood ratio, Wald, LM) will be discussed separately, with reference to a specific model type (e.g. LDVs in Weeks 11 to 13).

# Nonlinear regression: linear vs. nonlinear models

## Linear model:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

## Nonlinear model:

$$y_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i$$

- Linear model is a special case of the nonlinear model.
- Linear models are linear in parameters (encompass regressors such as  $x_i^2$ , etc.)
- Many nonlinear model may be transformed into linear models (log-transformation)
- For nonlinear models, nonlinear LS (NLS) are available.
- $\partial h(\mathbf{x}_i, \boldsymbol{\beta}) / \partial \mathbf{x}$  is no longer equal to  $\boldsymbol{\beta}$  (interpretation based on estimated model ...)

# Nonlinear regression

## Assumptions of the nonlinear regression model

- 1 **Functional form:** The conditional mean function for  $y_i$ , given  $\mathbf{x}_i$  is:

$$\mathbf{E}[y_i|\mathbf{x}_i] = h(\mathbf{x}_i, \boldsymbol{\beta}) , \quad i = 1, 2, \dots, n$$

- 2 **Identifiability of model parameters:** The parameter vector in the model is identified (estimable) if there is no nonzero parameter  $\boldsymbol{\beta}^0 \neq \boldsymbol{\beta}$  such that  $h(\mathbf{x}_i, \boldsymbol{\beta}^0) = h(\mathbf{x}_i, \boldsymbol{\beta})$  for all  $\mathbf{x}_i$ .
- 3 **Zero mean of the disturbance:** For  $y_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i$ , we assume

$$\mathbf{E}[\varepsilon_i|h(\mathbf{x}_i, \boldsymbol{\beta})] = 0 , \quad i = 1, 2, \dots, n$$

i.e. disturbance at observation  $i$  is uncorrelated with the conditional mean function.

# Nonlinear regression

## Assumptions of the nonlinear regression model

### 4 Homoskedasticity and nonautocorrelation:

conditional homoskedasticity:

$$\mathbf{E}[\varepsilon_i^2 | h(\mathbf{x}_i, \boldsymbol{\beta})] = \sigma^2, \quad i = 1, 2, \dots, n$$

nonautocorrelation:

$$\mathbf{E}[\varepsilon_i \varepsilon_j | h(\mathbf{x}_i, \boldsymbol{\beta}), h(\mathbf{x}_j, \boldsymbol{\beta})] = 0, \quad \text{for all } i \neq j$$

# Nonlinear regression

## Assumptions of the nonlinear regression model

- 5 **Data generating process:** DGP for  $\mathbf{x}_i$  is assumed to be a well-behaved population such that first and second sample moments of the data can be assumed to converge to fixed, finite population counterparts. The crucial assumption is that the process generating  $\mathbf{x}_i$  is strictly exogenous to that generating  $\varepsilon_i$
- 6 **Underlying probability model** There is a well-defined probability distribution generating  $\varepsilon_i$ . At this point, we assume only that this process produces a sample of uncorrelated, identically (marginally) distributed random variables  $\varepsilon_i$  with mean zero and variance  $\sigma^2$  conditioned on  $h(\mathbf{x}_i, \beta)$ . Thus, at this point, our statement of the model is **semi-parametric**.



# Nonlinear Regression: NLS

## NLS: estimator of the nonlinear regression model

- NLS:      min:     $S(\beta) = \sum [y_i - h(\mathbf{x}_i, \beta)]^2$
- Using standard procedure, we can get  $k$  first order conditions for the minimization:

$$\frac{\partial S(\beta)}{\partial \beta} = \sum_{i=1}^n [y_i - h(\mathbf{x}_i, \beta)] \frac{\partial h(\mathbf{x}_i, \beta)}{\partial \beta} = \mathbf{0}$$

- The above first order conditions are also moment conditions and this defines the NLS estimator as a GMM estimator.

# Nonlinear regression: NLS

## NLS: estimator of the nonlinear regression model

- NLS being a GMM estimator allows us to deduce that the NLS estimator has good large sample properties: consistency and asymptotic normality (if assumptions are fulfilled).
- Hypothesis testing: The principal testing procedure is the Wald test, which relies on the consistency and asymptotic normality of the estimator. Likelihood ratio and LM tests can also be constructed.

# Nonlinear regression: computing NLS estimates

For nonlinear models, a closed-form solution (NLS estimator) usually does not exist.

- Most of the nonlinear maximization problems are solved by an **iterative algorithm**.
- The most commonly used of iterative algorithms are **gradient methods**.
- The template for most gradient methods in common use is the **Newton's method**.
- Look at your software packages which methods are available for computing NLS estimates.

# Nonlinear regression: computing NLS estimates

Examples 7.4 & 7.8 (Greene):

Analysis of a Nonlinear Consumption Function

NLS with starting values equal to 0

NLS with starting values equal to the parameters from the OLS estimation (c(3) equal to 1).

Dependent Variable: REALCONS				
Method: Least Squares (Marquard - EViews legacy)				
Date: 09/19/16 Time 16:31				
Sample 1950Q1 2000Q4				
Included observations: 204				
REALCONS=C(1)+C(2)*REALDPI				
	Coefficient	Std.Error	t-Statistic	Prob.
C(1)	-80.35475	14.30585	-5.616915	0.0000
C(2)	0.921686	0.003872	238.0540	0.0000
R-squared	0.996448	Mean dependent var		2999.436
Adjusted R-squared	0.996431	S.D. dependent var		1459.707
S.E. of regression	87.20983	Akaike info criterion		11.78427
Sum squared resid	1536322	Schwarz criterion		11.81680
Log likelihood	-1199.995	Hannan-Quinn criter.		11.79743
F-statistics	56669.72	Durbin-Watson stat		0.092048
Prob(F-statistics)	0.000000			

# Nonlinear regression: computing NLS estimates

Examples 7.4 & 7.8 (Greene):

Analysis of a Nonlinear Consumption Function

---

Dependent Variable: REALCONS

Method: Least Squares (Marquard - EVIEWS legacy)

Sample 1950Q1 2000Q4    Included observations: 204

Convergence achieved after 200 iterations

REALCONS=C(1)+C(2)\*REALDPI^C(3)

---

	Coefficient	Std.Error	t-Statistic	Prob.
C(1)	458.7991	22.50140	20.38980	0.0000
C(2)	0.100852	0.010910	9.243667	0.0000
C(3)	1.244827	0.012055	103.2632	0.0000
R-squared	0.998834	Mean dependent var		2999.436
Adjusted R-squared	0.998822	S.D. dependent var		1459.707
S.E. of regression	50.09460	Akaike info criterion		10.68030
Sum squared resid	504403.2	Schwarz criterion		10.72910
Log likelihood	-1086.391	Hannan-Quinn criter.		10.70004
F-statistics	86081.29	Durbin-Watson stat		0.295995
Prob(F-statistics)	0.000000			

---

# Nonlinear regression: computing NLS estimates

Examples 7.4 & 7.8 (Greene):

Analysis of a Nonlinear Consumption Function

---

Dependent Variable: REALCONS

Method: Least Squares (Marquard - EVIEWS legacy)

Sample 1950Q1 2000Q4 Included observations: 204

Convergence achieved after 80 iterations

REALCONS=C(1)+C(2)\*REALDPI^C(3)

---

	Coefficient	Std.Error	t-Statistic	Prob.
C(1)	458.7989	22.50149	20.38971	0.0000
C(2)	0.100852	0.010911	9.243447	0.0000
C(3)	1.244827	0.012055	103.2632	0.0000
R-squared	0.998834	Mean dependent var		2999.436
Adjusted R-squared	0.998822	S.D. dependent var		1459.707
S.E. of regression	50.09460	Akaike info criterion		10.68030
Sum squared resid	504403.2	Schwarz criterion		10.72910
Log likelihood	-1086.391	Hannan-Quinn criter.		10.70004
F-statistics	86081.28	Durbin-Watson stat		0.295995
Prob(F-statistics)	0.000000			

---

## Quantile regression (QREG)

- Quantile regression estimates the relationship between regressors and a specified quantile of dependent variable.
- The (linear) quantile model can be defined as  $Q[y|\mathbf{x}, q] = \mathbf{x}'\boldsymbol{\beta}_q$ , such that  $\text{Prob}[y \leq \mathbf{x}'\boldsymbol{\beta}_q|\mathbf{x}] = q$ ,  $0 < q < 1$  where  $q$  denotes the  $q$ -th quantile.
- One important special case of quantile regression is the least absolute deviations (LAD) estimator, which corresponds to fitting the conditional median of the response variable ( $q = \frac{1}{2}$ ).
- QREG (LAD) estimator can be motivated as a robust alternative to OLS (with respect to outliers).

# Quantile regression (QREG)

For LRMs, the  $q$ -th quantile regression estimator  $\beta_q$  minimizes:

$$\min_{\hat{\beta}_q} Q_n(\hat{\beta}_q) = \sum_{i: e_i \geq 0}^n q|y_i - \mathbf{x}_i\hat{\beta}_q| + \sum_{i: e_i < 0}^n (1 - q)|y_i - \mathbf{x}_i\hat{\beta}_q|,$$

where  $e_i = (y_i - \mathbf{x}_i\hat{\beta}_q)$ .

- We use the notation  $\hat{\beta}_q$  to make clear that different choices of  $q$  lead to different  $\hat{\beta}$ .
- Slope of the loss function  $Q_n$  is asymmetrical (around  $e_i = 0$ ).
- The loss function is not differentiable (at  $e_i = 0$ )  
→ gradient methods are not applicable  
(linear programming can be used).



## Quantile regression - LAD

- LAD estimator is the QREG for  $q = \frac{1}{2}$  (median) and the loss function simplifies to:

$$\min_{\hat{\beta}_q} Q_n(\hat{\beta}_q) = \sum_{i=1}^n |y_i - \mathbf{x}_i \hat{\beta}_q|$$

- LAD estimator predates OLS (itself older than 200 years). Until recently, QREG and LAD have seen little use in econometrics, as OLS is vastly easier to compute.
- Different software packages use a variety of optimization algorithms for QREG/LAD estimation.
- Linear programming can be used for finding QREG estimates (Koenkerr and Bassett (around 1980)).

# Quantile regression

QREG coefficient interpretation example:

$$(1) \text{ wage}_i = \beta_0 + u_i$$

$$(2) \text{ wage}_i = \beta_0 + \beta_1 \text{female}_i + u_i$$

$$(3) \text{ wage}_i = \beta_0 + \beta_1 \text{female}_i + \beta_2 \text{exper}_i + u_i$$

The above equations are estimated by OLS / LAD / QREG:

Coefficient	OLS	LAD ( $q = \frac{1}{2}$ )	QREG ( $q = \frac{3}{4}$ )
(1) $\beta_0$	$\hat{\beta}_0 = \bar{y}$ sample mean	$\hat{\beta}_0 = \tilde{y}$ sample median	$\hat{\beta}_0 = Q_3$ sample 3 <sup>rd</sup> quartile
(2) $\beta_0, \beta_1$	conditional sample mean wage: male / female	conditional sample median wage: male / female	conditional sample $Q_3$ wage: male / female
(3) $\beta_2$	change in expected mean wage for $\Delta \text{exper} = 1$	change in expected median wage for $\Delta \text{exper} = 1$	change in expected $Q_3$ wage for $\Delta \text{exper} = 1$

# Quantile regression example

Example 7.9 (Greene):

Cobb-Douglass Production Function

OLS  $\rightarrow$  Standardized residuals indicate two outliers  $\rightarrow$  LAD

---

Dependent Variable: LN<sub>Y</sub>N

Method: Least Squares

Sample 1 25

Included observations: 25

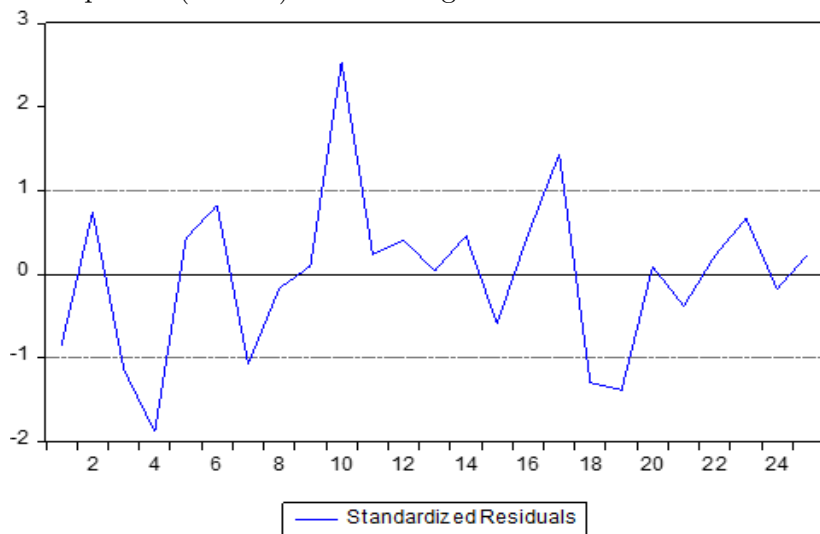
---

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.293263	0.107183	21.39582	0.0000
LN <sub>K</sub> N	0.278982	0.080686	3.457639	0.0022
LN <sub>L</sub> N	0.927312	0.012055	9.431359	0.0000
R-squared	0.959742	Mean dependent var		0.771734
Adjusted R-squared	0.956082	S.D. dependent var		0.899306
S.E. of regression	0.188463	Akaike info criterion		-0.387663
Sum squared resid	0.781403	Schwarz criterion		-0.241398
Log likelihood	7.845786	Hannan-Quinn criter.		-0.347095
F-statistics	262.2396	Durbin-Watson stat		1.937830
Prob(F-statistics)	0.000000			

---

# Quantile regression example

Example 7.9 (Greene): Cobb-Douglass Production Function



# Quantile regression example

Example 7.9 (Greene):

Cobb-Douglass Production Function  
(results differ from the textbook)

Dependent Variable: LNYN      Method: Quantile Regression (Median)				
Sample 1 25      Included observations: 25				
Huber Sandwich Standard Errors & Covariance				
Sparsity method: Kernel (Epanechnikov) using residuals				
Bandwidth method: Hall-Sheather, bw=0.33227				
Estimation successfully identifies unique optimal solution				
Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	2.275038	0.179268	12.69071	0.0000
LNKN	0.260365	0.122447	2.126351	0.0449
LNLN	0.927243	0.152593	6.076572	0.0000
Pseudo R-squared	0.794575	Mean dependent var		0.771734
Adjusted R-squared	0.775900	S.D. dependent var		0.899306
S.E. of regression	0.190505	Objective		1.627051
Quantile dependent va...	0.966677	Restr. objective		7.920415
Sparsity	0.594465	Quasi-LR statistic		84.69274
Prob(Quasi-LR stat)	0.000000			

# Quantile regression example 2

Example 7.10 (Greene):

Income Elasticity of Credit Cards Expenditure

OLS & LAD & Income Elasticity for 10 Deciles

---

Dependent Variable: LOGSPEND				
Method: Least Squares				
Date: 09/15/16 Time 13:53				
Sample (adjusted): 3 13443				
Included observations: 10499 after adjustments				

---

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	-3.055807	0.239699	-12.74852	0.0000
LOGINC	1.083438	0.032118	33.73296	0.0000
AGE	-0.017364	0.001348	-12.88069	0.0000
ADEPCNT	-0.044610	0.010921	-4.084857	0.0000

---

R-squared	0.100572	Mean dependent var	4.728778
Adjusted R-squared	0.100315	S.D. dependent var	1.404820
S.E. of regression	1.332496	Akaike info criterion	3.412366
Sum squared resid	18634.35	Schwarz criterion	3.415131
Log likelihood	-17909.21	Hannan-Quinn criter.	3.413300
F-statistic	391.1750	Durbin-Watson stat	1.888912
Prob(F-statistic)	0.000000		

---

# Quantile regression example 2

Example 7.10 (Greene):

Income Elasticity of Credit Cards Expenditure (LAD)

---

Dependent Variable: LOGSPEND    Method: Quantile Regression (Median)  
 Sample (adjusted): 3 13443    Included observations: 10499 after adjustments  
 Huber Sandwich Standard Errors & Covariance  
 Sparsity method: Kernel (Epanechnikov) using residuals  
 Bandwidth method: Hall-Sheather, bw=0.04437  
 Estimation successfully identifies unique optimal solution

---

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	-2.803756	0.233534	-12.00577	0.0000
LOGINC	1.074928	0.030923	34.76139	0.0000
AGE	-0.016988	0.001530	-11.10597	0.0000
ADEPCNT	-0.049955	0.011055	-4.518599	0.0000
Pseudo R-squared	0.058243	Mean dependent var		4.728778
Adjusted R-squared	0.057974	S.D. dependent var		1.404820
S.E. of regression	1.346476	Objective		5096.818
Quantile dependent va...	4.941583	Restr. objective		5412.032
Sparsity	2.659971	Quasi-LR statistic		948.0224
Prob(Quasi-LR stat)	0.000000			

---

# Quantile regression example 2

Example 7.10 (Greene):

Income Elasticity of Credit Cards Expenditure

