Advanced Econometrics

University of Warsaw

Jerzy Mycielski

Doctoral School of Social Sciences, 2024

Moment estimators

- 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 (OLS estimator)

- Replacing the variance of CEF error with sample variance:

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

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 $\widehat{S}(\beta)$

$$\widehat{\beta} = \mathrm{argmin}_{\beta \in R^k} \widehat{S}\left(\beta\right)$$

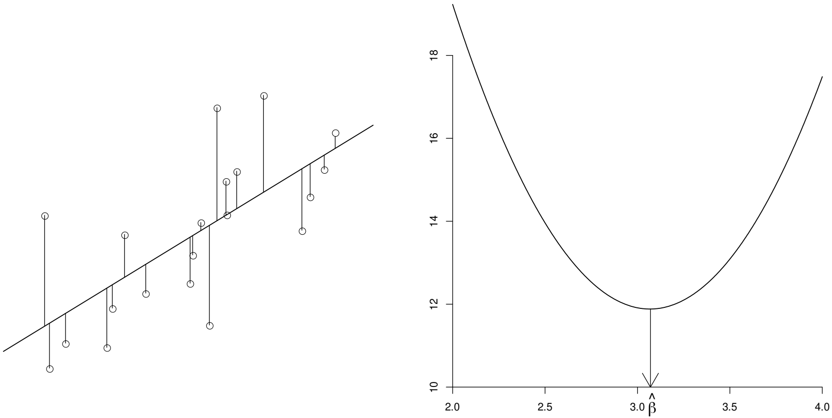
- First order conditions

$$\frac{\partial SSE\left(\beta\right)}{\partial \beta} = -2\sum_{i=1}^{n} X_i Y_i + 2\sum_{i=1}^{n} X_i X_i' \beta = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta = 0$$

- Least Squares Estimator is the solution of f.o.c.

$$\boxed{\widehat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}}$$

Sum of Squared Error, one regressor



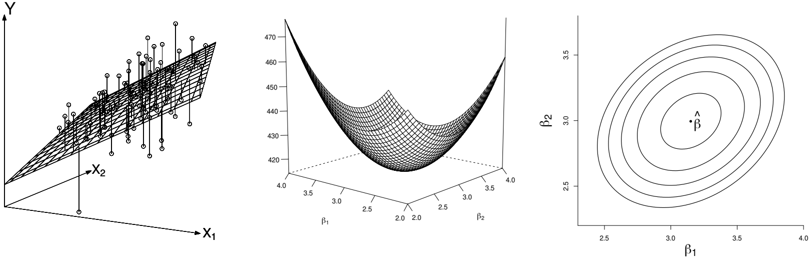
(a) Deviation from Fitted Line

(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)

Linear CEF estimation (OLS estimator)

- Alternative derivation of OLS estimator is based on replacing $\mathbf{Q}_{XX} = \mathbb{E}(XX')$ and $\mathbf{Q}_{XY} = \mathbb{E}(XY)$ by $\widehat{\mathbf{Q}}_{XX} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i'$ and $\widehat{\mathbf{Q}}_{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$ so that

$$\widehat{\beta} = \widehat{\mathbf{Q}}_{XX}^{-1}\widehat{\mathbf{Q}}_{XY}$$

- If matrix XX' is invertible $\widehat{\beta}$ unique. If XX' is not invertible we have multicollinearity problem.
- Notice that in this case XX' is positive definite and second order conditions for minimization of $SSE(\beta)$ are satisfied

$$\frac{\partial SSE\left(\widehat{\beta}\right)}{\partial \beta} = 2\mathbf{X}'\mathbf{X} > 0$$

where $\widehat{Y}_i = X_i'\widehat{\beta}$ is fitted value of Y_i from regression, $\widehat{e}_i = Y_i - \widehat{Y}_i$ is residual from regression
- Estimator of the error variance

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{e}_i^2$$

Analysis of variance, R^2

- It can be proven that

$$\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2=\sum_{i=1}^{n}\left(\widehat{Y}_i-\overline{Y}\right)^2+\sum_{i=1}^{n}e_i^2$$

- Coefficient of determination (R-squared)

$$R^2=\frac{\sum_{i=1}^{n}\left(\widehat{Y}_i-\overline{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2}$$

- Can be interpreted as the fraction of the sample variance of Y which is explained by the least squares fit.
- R^2 cannot be use used for comparing models as it always increases if variables are added to the model

Linear Regression Model

- The variables (Y, X) satisfy the linear regression equation

$$Y = X\beta + e.$$

$$\mathbb{E}\left[e | X \right] = 0$$

- Variables have finite second moments $\mathbb{E}\left[Y^2\right] < \infty$, $\mathbb{E}\left\|X\right\|^2 < \infty$ and an invertible design matrix $Q_{XX} = \mathbb{E}\left[XX'\right] > 0$.
- In addition above assumptions the homoskedasticity assumption is often made

$$\mathbb{E}\left(e^2 \middle| X \right) = \sigma^2 (X) = \sigma^2$$

- An estimator $\widehat{\theta}$ for θ is unbiased if $\mathbb{E}\left[\widehat{\theta}\right] = \theta$
- OLS estimator $\widehat{\beta}$ is unbiased if Linear Regression model is valid

Variance estimation

- Denote $\mathbf{D} = \mathrm{diag}\left(\sigma_1^2, ..., \sigma_n^2\right)$
- In the heteroscedastic linear regression model with i.i.d. sampling:

$$V_{\widehat{\beta}} = var\left(\widehat{\beta}\middle| X\right) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{D}\mathbf{X}\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}.$$

- If in addition the error is homoskedastic $\mathbf{V}_{\widehat{\beta}} = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$.
- Unbiased estimator of σ^2

$$s^2 = \frac{\sum_{i=1}^{n}\widehat{e}_i^2}{n-k}$$

- Unbiased estimator of $V_{\widehat{\beta}}$ for homoscedastic case

$$\mathbf{V}_{\widehat{\beta}}^0 = s^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

- Heteroskedasticity robust estimator of $V_{\widehat{\beta}}$

$$\boldsymbol{V}_{\widehat{\beta}}^{HC1} = \frac{n}{n-k}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\sum_{i=1}^{n}X_iX_i'\widehat{e}_i^2\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

Efficiency of the OLS estimator

Theorem

Gauss-Markov. In the homoscedastic linear regression model with i.i.d. sampling, if $\mathbb{E}\left(\widetilde{\beta}\middle|X\right)=\beta$ *then var* $\left(\widetilde{\beta}\middle|X\right)\geq\sigma^2\left(XX'\right)^{-1}$.

- OLS estimator is efficient!
- Assume either autocorrelation or heteroskedasticity is present. Then $var[e|X]=\Omega$.
- In such a case we can use GLS estimator:

$$\widehat{\beta}_{gls}=\left(\mathbf{X}'\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Y}.$$

- If $var[e|X]\neq\sigma^2 I$ then the GLS is efficient (not OLS).
- Modern practice is to use unefficient OLS estimator with heteroscedasticity robust variance matrix.
- However, care must be taken if sparse dummy are present in our data

Clustered samples

- Clustered sample model:

$$Y_{ig} = X_{ig}^{'}\beta + e_{ig} \text{ for } g = 1, \ldots, G; i = 1, \ldots, n_g$$

$$\mathbf{Y}_g = \mathbf{X}_g\beta + \mathbf{e}_g \text{ for } g = 1, \ldots, G$$

where $\mathbf{e}_g = (e_{1g}, ..., e_{n_g g}))$, g is the index of cluster, i is the index of observation inside the cluster

$$\widehat{\beta} = \left(\sum_{g=1}^{G}\mathbf{X}_g^{'}\mathbf{X}_g\right)^{-1}\left(\sum_{g=1}^{G}\mathbf{X}_g^{'}\mathbf{Y}_g\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

- We assume that erros for different clusters $(\mathbf{e}_g)$ and $(\mathbf{e}_h)$ are independent, but observations inside cluster can be dependent $var(\mathbf{e}_g) = \Omega$.

Clustered samples

- In the case of clustered samples we use clustered version of the variance estimator:

$$\widehat{\mathbf{V}}_\beta = a_n \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left(\sum_{g=1}^{G} \mathbf{X}_g^{'}\mathbf{e}_g\mathbf{e}_g^{'}\mathbf{X}_g\right) \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

$$a_n = \left(\frac{n-1}{n-k}\right)\left(\frac{G}{G-1}\right)$$

- The clustered estimate of the variance matrix is often very different from the standard unclustered version
- The number of clusters should be treated as being the number of observations
- It is very important to identify the correct level of clustering
  - if the clusters are too fine the variance matrix estimate has potentially large bias
  - if clusters are too large the estimated standard errors are very large

Clustered samples

- Duflo, Dupas and Kremer (2011) investigate the impact of tracking (assigning students based on initial test score)
- Clusters on the school level $G = 111$
- Estimate from the model

$$
\begin{array}{ccccc}
TestScore_{ig} = & -0.071 & + & 0.138 & Tracking_g + e_{ig}. \\
 & (0.019) & & (0.026) & \\
 & [0.054] & & 0.078] &
\end{array}
$$

Bibliography

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