Econometrics is …

the study of economics

using mathematical abstraction and statistical techniques

with data observations

An Example
Predicting 2016 Presidential Election

- To appreciate the use of econometrics to predict U. S. presidential elections, read:
  - Vote-Share Equations
  - Fair, R. C. (1996), "Econometrics and Presidential Elections," *Journal of Economic Perspective* 10, 89-102.
- Don't worry if you can not fully understand the econometric jargon used in the above readings. We will cover all of them this term, such as dummy variable and interaction variable, etc..

Classical Linear Regression Model

- Notation and Assumptions
- Model Estimation
  - Method of Moments
  - Least Squares
  - Partitioned Regression
- Model Interpretation

Notations

- $\mathbf{y}$: Dependent Variable (Regressand)
  - y_i, i = 1,2,…n
- $\mathbf{X}$: Explanatory Variables (Regressors)
  - $\mathbf{x}_i'$, i = 1,2,…,n
  - x_{ij}, i = 1,2,…,n; j = 1,2,…,K

Assumptions

- Assumption 1: Linearity
  $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + \varepsilon_i$
  $(i = 1,2,\ldots,n)$
- Linearity in Vector Notation
  $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$
  $(i = 1,2,\ldots,n)$

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

Assumptions

- Linearity in Matrix Notation
  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & \ldots & \vdots \\ x_{n1} & \cdots & x_{nK} \end{bmatrix}$$

Assumptions

- Assumption 2: Exogeneity
  $E(\varepsilon_i|\mathbf{X}) = E(\varepsilon_i|x_{j1},x_{j2},\ldots,x_{jK}) = 0$
  $(i,j = 1,2,\ldots,n)$
- Implications of Exogeneity
  - $E(\varepsilon_i) = E(E(\varepsilon_i|\mathbf{X})) = 0$
  - $E(x_{jk}\varepsilon_i) = E(E(x_{jk}\varepsilon_i|\mathbf{X})) = E(x_{jk}E(\varepsilon_i|\mathbf{X})) = 0$
  - $Cov(\varepsilon_i,x_{jk}) = E(x_{jk}\varepsilon_i) - E(x_{jk})E(\varepsilon_i) = 0$

Assumptions

- Assumption 3: No Multicolinearity
  rank($\mathbf{X}$) = K, with probability 1.
- Assumption 4: Spherical Error Variance
  $E(\varepsilon_i^2|\mathbf{X}) = \sigma^2 > 0$, i = 1,2,…,n
  $E(\varepsilon_i\varepsilon_j|\mathbf{X}) = 0$, i,j = 1,2,…,n; $i \neq j$
  In matrix notation:
  $Var(\boldsymbol{\varepsilon}|\mathbf{X}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}) = \sigma^2\mathbf{I}_n$

Assumptions

- Implication of Spherical Error Variance (Homoscedasticity and No Autocorrelation)
  - $\mathrm{Var}(\boldsymbol{\varepsilon}) = \mathrm{E}(\mathrm{Var}(\boldsymbol{\varepsilon}|\mathbf{X}))+\mathrm{Var}(\mathrm{E}(\boldsymbol{\varepsilon}|\mathbf{X})) = \mathrm{E}(\mathrm{Var}(\boldsymbol{\varepsilon}|\mathbf{X})) = \sigma^2\mathbf{I}_n$
  - $\mathrm{Var}(\mathbf{X}'\boldsymbol{\varepsilon}|\mathbf{X}) = \mathrm{E}(\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}|\mathbf{X}) = \sigma^2\mathrm{E}(\mathbf{X}'\mathbf{X})$
    Note: $\mathrm{E}(\mathbf{X}'\boldsymbol{\varepsilon}|\mathbf{X}) = 0$ followed from Exogeneity Assumption

Discussions

- Exogeneity in Time Series Models
- Random Samples
  The sample (**y**,**X**) is a random sample if $\{y_i, \mathbf{x}_i\}$ is i.i.d (independently and identically distributed).
- Fixed Regressors
  X is fixed or deterministic.

Discussions

- Nonlinearity in Variables
  $g(y_i) = \beta_1 f_1(x_{i1})+\beta_2 f_2(x_{i2})+\ldots+ \beta_K f_K(x_{iK})+\varepsilon_i$
  $(i = 1,2,\ldots,n)$
  - Linearity in Parameters and Model Errors

$$y_i = e^{\mathbf{x}_i^{'}\boldsymbol{\beta}} e^{\varepsilon_i} = e^{\mathbf{x}_i^{'}\boldsymbol{\beta}+\varepsilon_i} \Rightarrow \ln(y_i) = \mathbf{x}_i^{'}\boldsymbol{\beta} + \varepsilon_i$$

$$y_i = \alpha \prod_{k=1}^{K} x_{ik}^{\beta_k} e^{\varepsilon_i} \Rightarrow \ln(y_i) = \beta_0 + \sum_{k=1}^{K} \ln(x_{ik})\beta_k + \varepsilon_i$$

Method-of-Moments Estimator

- From the implication of strict exogeneity assumption, $E(x_{jk}\varepsilon_i) = 0$ for i = 1,2,…,n and k = 1,2,…,K. That is,
- Moment Conditions: $E(\mathbf{X}'\boldsymbol{\varepsilon}) = \mathbf{0}$
- $\mathbf{X}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$ or $\mathbf{X}'\mathbf{y}=\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$
- The method-of-moments estimator of $\boldsymbol{\beta}$: $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

Least Squares Estimator

- OLS: $\mathbf{b} = \text{argmin}_{\beta}\ SSR(\boldsymbol{\beta})$
- $SSR(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$
  $= (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$
  $= \sum_{i=1,2\ldots,n}(y_i-\mathbf{x}_i'\boldsymbol{\beta})^2$
- $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

The Algebra of Least Squares

- $$\frac{\partial \mathrm{SSR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

  $$\frac{\partial^2 \mathrm{SSR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = 2\mathbf{X}'\mathbf{X}$$

- Normal Equations: $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$
- $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

The Algebra of Least Squares

- $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$
- $\mathbf{b} = (\mathbf{X}'\mathbf{X}/n)^{-1}(\mathbf{X}'\mathbf{y}/n) = \mathbf{S}_{xx}^{-1}\mathbf{s}_{xy}$
  $\mathbf{S}_{xx}$ is the sample average of $\mathbf{x}_i\mathbf{x}_i'$, and $\mathbf{s}_{xy}$ is the sample average of $\mathbf{x}_i y_i$.
- OLS Residuals: $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$
  $\mathbf{X}'\mathbf{e} = \mathbf{0}$
- $s^2 = SSR/(n-K) = \mathbf{e}'\mathbf{e}/(n-K)$

The Algebra of Least Squares

- Projection Matrix **P** and "Residual Maker" **M**:
  - $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{M} = \mathbf{I}_n - \mathbf{P}$
  - $\mathbf{PX} = \mathbf{X}$, $\mathbf{MX} = \mathbf{0}$
  - $\mathbf{Py} = \mathbf{Xb}$, $\mathbf{My} = \mathbf{y\text{-}Py} = \mathbf{e}$, $\mathbf{M}\boldsymbol{\varepsilon} = \mathbf{e}$
- SSR = $\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$
- **h** = Vector of diagonal elements in **P** can be used to check for the **influential observation**: $h_i > K/n$, with $0 \leq h_i \leq 1$.

The Algebra of Least Squares

- Fitted Value: $\hat{\mathbf{y}} = \mathbf{Xb} = \mathbf{Py}, \quad \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$
- Uncentered R^2: $R_{uc}^2 = 1 - \mathbf{e'e}/\mathbf{y'y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} / \mathbf{y'y}$
- Centered R^2: $R^2 = 1 - \mathbf{e'e}/(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}})$

$$(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}}) = (\hat{\mathbf{y}} - \bar{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}}) + \mathbf{e'e}$$

with $\bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Analysis of Variance

- TSS = $(\mathbf{y}-\bar{\mathbf{y}})'(\mathbf{y}-\bar{\mathbf{y}})$
- ESS = $(\hat{\mathbf{y}}-\bar{\mathbf{y}})'(\hat{\mathbf{y}}-\bar{\mathbf{y}})$, where $\bar{\mathbf{y}} = \bar{\hat{\mathbf{y}}}$
- RSS = $\mathbf{e}'\mathbf{e}$ (or SSR)
- TSS = ESS + RSS
- Degrees of Freedom
  - TSS: n-1
  - ESS: K-1
  - RSS: n-K

Analysis of Variance

- R^2

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\mathbf{e}'\mathbf{e}}{(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}})}$$

- Adjusted R^2

$$\bar{R}^2 = 1 - \frac{\frac{RSS}{n-K}}{\frac{TSS}{n-1}} = 1 - \frac{\frac{\mathbf{e}'\mathbf{e}}{n-K}}{\frac{(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}})}{n-1}}$$

Discussions

- Examples
  - Constant or Mean Regression

    $y_i = \alpha + \varepsilon_i$ (i=1,2,…,n)

    $$a = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

  - Simple Regression

    $y_i = \alpha + \beta x_i + \varepsilon_i$ (i=1,2,…,n)

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad a = \bar{y} - b\bar{x}$$

Partitioned Regression

- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}1\boldsymbol{\beta}1 + \mathbf{X}2\boldsymbol{\beta}2 + \boldsymbol{\varepsilon}$
  $\mathbf{X} = [\mathbf{X}1\ \mathbf{X}2],\ \boldsymbol{\beta} = [\boldsymbol{\beta}1'\ \ \boldsymbol{\beta}2']'$
- Normal Equations
  $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$, or

$$\begin{bmatrix} \mathbf{X}1'\mathbf{X}1 & \mathbf{X}1'\mathbf{X}2 \\ \mathbf{X}2'\mathbf{X}1 & \mathbf{X}2'\mathbf{X}2 \end{bmatrix} \begin{bmatrix} \mathbf{b}1 \\ \mathbf{b}2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}1'\mathbf{y} \\ \mathbf{X}2'\mathbf{y} \end{bmatrix}$$

Partitioned Regression

- $\mathbf{b}1 = (\mathbf{X}1'\mathbf{X}1)^{-1}\,\mathbf{X}1'\mathbf{y} - (\mathbf{X}1'\mathbf{X}1)^{-1}(\mathbf{X}1'\mathbf{X}2)\mathbf{b}2$
  $= (\mathbf{X}1'\mathbf{X}1)^{-1}\mathbf{X}1'(\mathbf{y}-\mathbf{X}2\mathbf{b}2)$
- $\mathbf{b}2 = (\mathbf{X}2'\mathbf{M}1\mathbf{X}2)^{-1}\,\mathbf{X}2'\mathbf{M}1\mathbf{y} = (\mathbf{X}2^{*\prime}\mathbf{X}2^{*})^{-1}\mathbf{X}2^{*\prime}\mathbf{y}^{*}$
  $\mathbf{M}1 = [\mathbf{I}-\mathbf{X}1(\mathbf{X}1'\mathbf{X}1)^{-1}\mathbf{X}1']$
  $\mathbf{X}2^{*} = \mathbf{M}1\mathbf{X}2$ (matrix of residuals obtained from each column of **X**2 regressed on **X**1)
  $\mathbf{y}^{*} = \mathbf{M}1\mathbf{y}$ (residuals of **y** regressed on **X**1)
- **b**2 is LS estimator of $\mathbf{y}^{*}$ on $\mathbf{X}2^{*}$

Partitioned Regression

- **Frisch-Waugh Theorem**
  If $\mathbf{y} = \mathbf{X}1\boldsymbol{\beta}1 + \mathbf{X}2\boldsymbol{\beta}2 + \boldsymbol{\varepsilon}$, then $\mathbf{b}2$ is the LS estimator for the regression of residuals *from a regression of **y** on **X**1 alone* are regressed on the set of residuals *obtained when each column of **X**2 is regressed on **X**1*.
- If $\mathbf{X}1$ and $\mathbf{X}2$ are independent, $\mathbf{X}1'\mathbf{X}2 = 0$.
  $\mathbf{b}1 = (\mathbf{X}1'\mathbf{X}1)^{-1}\mathbf{X}1'\mathbf{y}$
  $\mathbf{b}2 = (\mathbf{X}2'\mathbf{X}2)^{-1}\mathbf{X}2'\mathbf{y}$

Partitioned Regression

- **Applications of Frisch-Waugh Theorem**
  - Partial Regression Coefficient
  - Trend Regression

    $y_t = \alpha + \beta t + \varepsilon_t$

Model Interpretation

- Marginal Effects (*Ceteris Paribus* Interpretation): **y = Xb + e**, where $b_k = \frac{\partial \mathbf{y}}{\partial \mathbf{x}_k}$
- Elasticity Interpretation

$$y_i = e^{\mathbf{x}_i^{'}\boldsymbol{\beta}} e^{\varepsilon_i} = e^{\mathbf{x}_i^{'}\boldsymbol{\beta}+\varepsilon_i} \Rightarrow \ln(y_i) = \mathbf{x}_i^{'}\boldsymbol{\beta} + \varepsilon_i \Rightarrow b_k = \frac{\partial \ln(\mathbf{y})}{\partial \mathbf{x}_k} = \frac{1}{\mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}_k}$$

$$y_i = \alpha \prod_{k=1}^{K} x_{ik}^{\beta_k} e^{\varepsilon_i} \Rightarrow \ln(y_i) = \beta_0 + \sum_{k=1}^{K} \ln(x_{ik})\beta_k + \varepsilon_i$$

$$b_k = \frac{\partial \ln(\mathbf{y})}{\partial \ln(\mathbf{x}_k)} = \frac{\mathbf{x}_k}{\mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}_k}$$

Example

- U. S. Gasoline Market, 1953-2004
  - EXPG = Total U.S. gasoline expenditure
  - PG = Price index for gasoline
  - Y = Per capita disposable income
  - Pnc = Price index for new cars
  - Puc = Price index for used cars
  - Ppt = Price index for public transportation
  - Pd = Aggregate price index for consumer durables
  - Pn = Aggregate price index for consumer nondurables
  - Ps = Aggregate price index for consumer services
  - Pop = U.S. total population in thousands

Example

- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
- $\mathbf{y}$ = G; $\mathbf{X}$ = [1 PG Y]
  where G = (EXPG/PG)/POP
- $\mathbf{y}$ = *ln*(G); $\mathbf{X}$ = [1 *ln*(PG) *ln*(Y)]
- Ex: $\mathbf{X}$ = [1 *ln*(PG) *ln*(Y) *ln*(Pnc) *ln*(Puc)]
- Elasticity Interpretation

Example (Continued)

- $\mathbf{y} = \mathbf{X}1\boldsymbol{\beta}1 + \mathbf{X}2\boldsymbol{\beta}2 + \boldsymbol{\varepsilon}$
- $\mathbf{y} = \mathit{ln}(G)$; $\mathbf{X1} = [1 \text{ YEAR}]$;
  $\mathbf{X}2 = [\mathit{ln}(PG)\ \mathit{ln}(Y)]$
  where G = (EXPG/PG)/POP
- **Frisch-Waugh Theorem**