ECO375: Review Notes

Applied Econometrics I

Tianyu Du

October 27, 2018

Contents

1 Slide 4: Simple & Multiple Regression - Estimation

1.1 Regression Model

Assumption 1.1. Assuming the population follows

$$y = \beta_0 + \beta_1 x + u$$

and assume that x *causes* y.

1.2 OLS

$$\min_{\vec{\beta}} \sum_i (y_i - \hat{y}_i)^2$$

With FOC:

$$\sum_i (y_i - \hat{y}_i) = 0$$

$$\sum_i x_{ij}(y_i - \hat{y}_i) = 0, \ \forall j$$

Remark 1.1. Both $\hat{\beta}_0$ and $\hat{\beta}_j$ are functions of *random variables* and therefore themselves *random* with *sampling distribution*. And the estimated coefficients are random up to random sample chosen.

Property 1.1. Properties of OLS estimators

- **Unbiased** $\mathbb{E}[\hat{\beta}|X] = \beta$
- **Consistent** $\hat{\beta} \to \beta$ as $n \to \infty$
- **Efficient/Good** min variance.

Definition 1.1. The **Simple Coefficient of Determination**

$$R^2 = \frac{SSE}{SST}$$

and $SS\underline{Total} = SS\underline{Explained} + SS\underline{Residual}$

$$\sum_i (y_i - \overline{y})^2 = \sum_i (\hat{y}_i - \overline{y})^2 + \sum_i (y_i - \hat{y}_i)^2$$

Proposition 1.1 (Logarithms)**.** Interpretation with logarithmic transformation.

- $\ln y = \alpha + \beta \ln y + u$: $\underline{x \text{ increases by } 1\%, \; y \text{ increases by } \beta\%}$.
- $\ln y = \alpha + \beta x + u$: $\underline{x \text{ increases by 1 unit}, \; y \text{ increases by } 100\beta\%}$.
- $y = \alpha + \beta \ln x + u$: $\underline{x \text{ increases by } 1\%, \; y \text{ increases by } 0.01\beta \text{ unit.}}$

Assumption 1.2. Simple regression model assumptions

1. Model is <u>linear</u> in parameter.
2. <u>Random samples</u> $\{(x_i, y_i)\}_{i=1}^n$.
3. Sample outcomes $\{x_i\}_{i=1}^n$ are not the same.
4. $\mathbb{E}(u|x) = 0$ conditional on random sample x.
5. Error is <u>homoskedastic</u>. $Var(u|x) = \sigma^2$ for all x.

Benefits of MLR compared with SLR

- More accurate causal effect estimation.
- More flexible function forms.
- Could explicitly include more predictors so $\mathbb{E}(u|X) = 0$ is easier to be satisfied.
- MLR4 is less restrictive than SLR4.

Property 1.2. MLR OLS residual satisfies

$$\sum_i \hat{u}_i = 0$$

$$\sum_i x_{ji} \hat{u}_i = 0, \; \forall i \in \{1, 2, \ldots, k\}$$

Property 1.3. MLR OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ pass through the average point.

$$\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}_1 + \cdots + \hat{\beta}_k \overline{x}_k$$

Proof. ■

1.3 Partialling Out

1.3.1 Steps

1. Regress x_1 on $x_2, x_3, \ldots, x_K$ and calculate the residual $\widetilde{r}_1$.
2. Regress y on $\widetilde{r}_1$ with simple regression and find the estimated coefficient $\hat{\lambda}_1$.
3. Then the multiple regression coefficient estimator $\hat{\beta}_1$ is

$$\hat{\beta}_1 = \hat{\lambda}_1 = \frac{\sum_i y_i \widetilde{r}_{1i}}{\sum_i (\widetilde{r}_{1i})^2}$$

Proof. ■

1.3.2 Interpretation

This OLS estimator only uses the unique variance of one independent variable. And the parts of variation correlated with other independent variables is partialled out.

Assumption 1.3. Multiple Regression Assumptions

1. (MLR1) The model is linear in parameters.
2. (MLR2) Random sample from population $\{(x_{1i}, \ldots x_{ki}, y_i\}_{i=1}^n$.
3. (MLR3) No perfect multicollinearity.
4. (MLR4) Zero expected error conditional on population slice given by X.

$$\mathbb{E}(u|X) = \mathbb{E}(u|x_1, x_2, \ldots, x_k) = 0$$

5. (MLR5) Homoskedastic error conditional on population slice given by X.

$$Var(u|X) = \sigma^2$$

6. (MLR6, *strict assumption*) Normally distributed error

$$u \sim \mathcal{N}(0, \sigma^2)$$

1.4 Omitted Variable Bias

Suppose population follows the *real model*

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i \tag{1}$$

Consider the *alternative model*, and x_k is omitted, which is assumed to be relevant.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{k-1} x_{(k-1)i} + r_i \tag{2}$$

and use the partialling-out result on the second regression we have

$$\tilde{\beta}_1 = \frac{\sum_i \tilde{r}_{1i} y_i}{(\tilde{r}_{1i})^2}$$

where $\tilde{r_1}_i = x_{1i} - \tilde{\alpha}_0 - \tilde{\alpha}_2 x_{2i} - \cdots - \tilde{\alpha}_{k-1} x_{(k-1)i}$

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_k \frac{\sum(\tilde{r}_{1i} x_{ki})}{\sum(\tilde{r}_{1i})^2} \tag{3}$$

and take the expectation

$$\mathbb{E}(\tilde{\beta}_1 | X) = \beta_1 + \tilde{\delta}_1 \beta_k$$
$$Bias(\tilde{\beta}_1) = \tilde{\delta}_1 \beta_k$$

Conclusion the sign of bias depends on $cov(x_1, x_k)$ and β_k.

Proof. TODO ∎

2 Matrix Differentiation*

$$\mathbf{y} = \mathbf{A}\mathbf{x} \implies \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A} \tag{4}$$

Let $\alpha = \mathbf{y}'\mathbf{A}\mathbf{x}$, notice that $\alpha \in \mathbb{R}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y}'\mathbf{A} \tag{5}$$
$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}'\mathbf{A}' \tag{6}$$

Consider special case $\alpha = \mathbf{x}'\mathbf{A}\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}'\mathbf{A} + \mathbf{x}'\mathbf{A}' \tag{7}$$

and if $\mathbf{A}$ is symmetric,

$$\frac{\partial \alpha}{\partial \mathbf{x}} = 2\mathbf{x}'\mathbf{A} \tag{8}$$

3 Multiple Regression in Matrices

3.1 The Model

Predictor

$$\mathbf{X} \in \mathbb{M}_{n \times (k+1)}(\mathbb{R})$$

where n is the number of observations and k is the number of features.

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & & & \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}_{n \times (k+1)}$$

Model

$$\mathbf{y} = \mathbf{X}\vec{\beta} + \mathbf{u}$$

First order condition for OLS

$$\begin{aligned} \mathbf{X}'\hat{u} &= \mathbf{0} \in \mathbb{R}^{k+1} \\ \iff \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) &= \mathbf{0} \in \mathbb{R}^{k+1} \end{aligned}$$

Estimator

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Proof. From the first order condition for the OLS estimator

$$\begin{aligned} \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) &= \mathbf{0} \\ \implies \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\beta} &= \mathbf{0} \\ \implies \mathbf{X}'\mathbf{y} &= \mathbf{X}'\mathbf{X}\hat{\beta} \\ \implies \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \end{aligned}$$

and note that $(\mathbf{X}'\mathbf{X})$ is guaranteed to be invertible by assumption *no perfect multi-collinearity.* ■

Sum Squared Residual

$$SSR(\hat{\beta}) = \hat{u}' \cdot \hat{u} = (\mathbf{y} - \mathbf{X}\hat{\beta})' \cdot (\mathbf{y} - \mathbf{X}\hat{\beta})$$

3.2 Variance Matrix

Consider

$$\begin{aligned} \vec{z}_t &= [z_{1t}, z_{2t}, \dots z_{nt}]' \\ \vec{z}_s &= [z_{1s}, z_{2s}, \dots z_{ns}]' \end{aligned}$$

Notice that the variance and covariance are defined as

$$\begin{aligned} Var(\vec{z}_t) &= \mathbb{E}[(\vec{z}_t - \mathbb{E}[\vec{z}_t])^2] \\ Cov(\vec{z}_t, \vec{z}_s) &= \mathbb{E}[(\vec{z}_t - \mathbb{E}[\vec{z}_t])(\vec{z}_s - \mathbb{E}[\vec{z}_s])] \end{aligned}$$

The **variance matrix** of $\mathbf{z} = [z_1, z_2, \dots, z_n]$ is given by

$$Var(\mathbf{z}) = \begin{bmatrix} Var(z_1) & Cov(z_1, z_2) & \dots & Cov(z_1, z_n) \\ Cov(z_2, z_1) & \dots & & \\ \vdots & & & \\ Cov(z_n, z_1) & \dots & \dots & Var(z_n) \end{bmatrix}$$

$$= \begin{bmatrix} \mathbb{E}[(z_1 - \overline{z}_1)^2] & \mathbb{E}[(z_1 - \overline{z}_1)(z_2 - \overline{z}_2)] & \dots \\ \mathbb{E}[(z_2 - \overline{z}_2)(z_1 - \overline{z}_1)] & \dots & \\ \vdots & & \\ \mathbb{E}[(z_n - \overline{z}_n)(z_1 - \overline{z}_1)] & \dots & \mathbb{E}[(z_n - \overline{z}_n)^2] \end{bmatrix}$$

$$= \mathbb{E}[(\mathbf{z} - \mathbb{E}[\mathbf{z}])_{n\times 1} \cdot (\mathbf{z} - \mathbb{E}[\mathbf{z}])'_{1\times n}] \in \mathbb{M}_{n\times n}$$

In the special case $\mathbb{E}[\vec{z}] = \vec{0}$, variance is reduced to

$$Var(\mathbf{z}) = \mathbb{E}[\mathbf{z} \cdot \mathbf{z}']$$

Residual Since residual u_i are $i.i.d$ with variance σ^2, the variance matrix of $\mathbf{u}$ is

$$Var(\mathbf{u}) = \mathbb{E}[\mathbf{u} \cdot \mathbf{u}'] = \sigma^2 \mathbf{I}_n$$

Estimator If $\hat{\beta}$ is unbiased, $\mathbb{E}[\hat{\beta}|\mathbf{X}] = \vec{\beta}$, then

$$Var(\hat{\beta}|\mathbf{X}) = \mathbb{E}[(\hat{\beta} - \vec{\beta}) \cdot (\hat{\beta} - \vec{\beta})'|\mathbf{X}] \in \mathbb{M}_{(k+1)\times(k+1)}$$

4 Slide 7

4.1 Assumptions (MLRs) in Matrix Form

E.1. *linear in parameter*

$$\mathbf{y} = \mathbf{X}\vec{\beta} + \mathbf{u}$$

E.2. *no perfect multi-collinearity*

$$rank(\mathbf{X}) = k + 1$$

E.3. Error has expected value of $\mathbf{0}$ conditional on $\mathbf{X}$.

$$\mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbf{0}$$

E.4. Error $\mathbf{u}$ is *homoscedastic*.

$$Var(\mathbf{u}|\mathbf{X}) = \sigma^2 \mathbf{I}_n$$

E.5. *Normally distributed* error $\mathbf{u}$. Note that this assumption is relatively strong.

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

4.2 Properties of OLS Estimator

Theorem 4.1. Given *E.1. E.2. E.3.*, the OLS estimator $\hat{\beta}$ is an unbiased estimator for $\vec{\beta}$.

$$\mathbb{E}[\hat{\beta}|\mathbf{X}] = \vec{\beta}$$

Proof.

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\vec{\beta} + \mathbf{u}) \\
&= \vec{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}
\end{aligned}$$

Taking expectation conditional on $\mathbf{X}$ on both sides,

$$\mathbb{E}[\hat{\beta}|\mathbf{X}] = \vec{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} = \vec{\beta}$$

■

Lemma 4.1. Suppose $\mathbf{A} \in \mathbb{M}_{m\times n}$ and $\mathbf{z} \in \mathbb{M}_{n\times 1}$ then

$$Var(\mathbf{Az}) = \mathbf{A}Var(\mathbf{z})\mathbf{A}'$$

Theorem 4.2. Given $E.1 \sim E.4$

$$Var(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

Proof.

$$\begin{aligned}
Var(\hat{\beta}|\mathbf{X}) &= Var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}|\mathbf{X}) \\
&= Var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\vec{\beta} + \mathbf{u})|\mathbf{X}) \\
&= Var(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X})
\end{aligned}$$

By the lemma above,

$$\begin{aligned}
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Var(\mathbf{u}|\mathbf{X})[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']' \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Var(\mathbf{u}|\mathbf{X})\mathbf{X}''(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}_n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

■

Theorem 4.3 (Gause-Markov)**.** Given $E.1. \sim E.4.$, the OLS estimator is the best linear unbiased estimator(BLUE).
(*The best* here means the OLS has the least variance among all estimators.)

4.3 Variance Inflation

Let $j \in \{1, 2, \dots, k\}$, then the variance of an individual estimator on particular feature j is

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{(1 - R_j^2)SST_j}$$

where

$$SST_j = \sum_{i=1}^{n} (x_{ij} - \overline{x}_j)^2$$

and R_j^2 is the coefficient of determination while regressing x_j on <u>all other</u> features $x_i, \forall i \neq j$.

Definition 4.1. The **variance inflation** on estimator for feature j is

$$VIF_j = \frac{1}{1 - R_j^2}$$

Remark 4.1 (Interpretation)**.** the standard error of estimator on a particular variable $(\hat{\beta}_j)$ is *inflated* by it's(x_j) relationship with other explanatory variables.

Solutions to high VIF

1. Drop the explanatory variable.
2. Use ratio $\frac{x_i}{x_j}$ instead.
3. Ridge regression.

Remark 4.2. VIF highlights the importantce of **not** including redundant predictors.

5 Slide 8: Multiple Regression-Inference

Hypothesis Testing on multiple regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots \beta_k x_{ik} + u_i$$

5.1 t-test for significance of individual predicator

Test statistic Given $MLR.1 \sim MLR.6$ (need $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$),

$$t = \frac{\hat{\beta}_j - b}{s.e.(\hat{\beta}_j)} \sim t_{n-k-1}$$

where

$$H_0 : \beta_j = b$$
$$H_1 : \beta_j (\neq, >, <) b$$

5.2 t-test for comparing 2 coefficients

Test statistic

$$t = \frac{(\hat{\beta}_i - \hat{\beta}_j) - b}{s.e.(\hat{\beta}_i - \hat{\beta}_j)} \sim t_{n-k-1}$$

where

$$H_0 : \beta_i - \beta_j = b$$
$$H_1 : \beta_i - \beta_j (\neq, >, <) b$$

notice

$$s.e.(\hat{\beta}_i - \hat{\beta}_j) = \sqrt{Var(\hat{\beta}_i - \hat{\beta}_j)}$$
$$= \sqrt{Var(\hat{\beta}_i) + Var(\hat{\beta}_j) - 2Cov(\hat{\beta}_i, \hat{\beta}_j)}$$

5.3 Partial F-test for joint significance

$$H_0 : \beta_i = \beta_j = \beta_k = \cdots = 0$$
$$H_1 : \exists\ z \in \{i, j, k, \dots\}\ s.t.\ \beta_z \neq 0$$

Test significance by comparing the *restricted* and *unrestricted* models, see whether restricting the model by removing certain explanatory variables "significantly" hurts the fit of the model.

$$df = (q, n-k-1)$$

Test statistic

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \sim F_{(q,n-k-1)}$$

or

$$F' = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)} \sim F_{(q,n-k-1)}$$

5.4 Full F-test for the significance of the model

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$
$$H_1 : \exists\ i \in \{1, 2, \dots, 3\}\ s.t.\ \beta_i \neq 0$$

Remark 5.1. R^2 version only and substitute $R^2_r = 0$, since SSR_r is undefined.

Test statistic

$$F = \frac{R^2_{ur}/k}{(1 - R^2_{ur})/(n-k-1)} \sim F_{(k,n-k-1)}$$

5.5 F-test for general restrictions

Remark 5.2. Use the SSR version of $Fstatistic$ only since the SST for restricted and unrestricted models are different.

Remark 5.3. We only reject or failed to reject H_0, we never accept H_0 in a hypothesis test.

6 Slides 9

6.1 Data Scaling

6.1.1 Mutiplier

1. Enlarge x_j by factor a: $\hat{\beta}_j$ shrinks by a.
2. Enlarge y by factor a: **all** $\hat{\beta}_i$ enlarged by a.
3. Test statistic $t = \frac{\hat{\beta}}{s.e.(\hat{\beta})} = \frac{a\hat{\beta}}{s.e.(a\hat{\beta})}$ is unaffected.

6.1.2 Standardization

Standardized variable For j^{th} observation of explanatory variable x,

$$z_j = \frac{x_j - \overline{x}}{\sigma_x}$$

which satisfies

$$\mathbb{E}[z_j] = 0,\ Var(z_j) = 1$$

Properties Consider model and find the estimator of regressing standardized y on standardized x.

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik} + \hat{u}_i$$

Since OLS estimator passes through the mean,

$$\begin{aligned}
\overline{y} &= \hat{\beta}_0 + \hat{\beta}_1\overline{x}_1 + \dots \hat{\beta}_k\overline{x}_k \\
\implies (y_i - \overline{y}) &= \hat{\beta}_1(x_{i1} - \overline{x}_1) + \cdots + \hat{\beta}_k(x_{ik} - \overline{x}_k) + \hat{u}_i \\
\implies \frac{y_i - \overline{y}}{\sigma_y} &= \frac{\hat{\beta}_1\sigma_{x_1}}{\sigma_y}\frac{x_{i1} - \overline{x}_1}{\sigma_{x_1}} + \cdots + \frac{\hat{\beta}_k\sigma_{x_k}}{\sigma_y}\frac{x_{ik} - \overline{x}_k}{\sigma_{x_k}} + \frac{\hat{u}_i}{\sigma_y} \\
\implies b_j &= \frac{\hat{\beta}_j\sigma_{x_j}}{\sigma_y}
\end{aligned}$$

Remark 6.1 (Interpretation)**.** x_j increases by 1 **std**, y increases by $b_j = \frac{\hat{\beta}_j\sigma_{x_j}}{\sigma_y}$ **std**, *ceteris paribus.*

6.2 Logarithmic Function

Exact interpretation of log transformation.

$$\ln(y_i) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots \hat{\beta}_k x_{ik} + \hat{u}_i$$

Derive.

$$\begin{aligned}
\ln(y_2) - \ln(y_1) &= \hat{\beta}_j \Delta x_j \\
\implies \ln(\frac{y_2}{y_1}) &= \hat{\beta}_j \Delta x_j \\
\implies \frac{y_2}{y_1} &= exp(\hat{\beta}_j \Delta x_j) \\
\implies \frac{y_2 - y_1}{y_1} &= \frac{y_2}{y_1} - 1 \\
\implies \%\Delta y &= exp(\hat{\beta}_j \Delta x_j) - 1
\end{aligned}$$

■

6.3 Quadratics and Polynomials

Model

$$y_i = \sum_{p=0}^{k} \beta_p x_i^p + u_i$$

Remark 6.2. Consider the **interpretation** and **turning points**.

6.4 Interaction Effects

Consider model

$$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz + u$$

then

$$\frac{\partial y}{\partial x} = \beta_1 + \beta_3 z$$

1. The effects of change of x on y depends on z.
2. Interpretation: *evaluate* $\frac{\partial y}{\partial x}$ at a z point that we are interested in.
3. Use *conventional testing* (t-test) to check if interaction term is significant.

6.5 Regression Selection and Adjusted R-square

The adjusted R-square, $\overline{R^2}$, incorporates a *penalty* for including more regressors (if insignificant).

$$\overline{R^2} = 1 - \frac{(1 - R^2)(n-1)}{n-k-1}$$

Remark 6.3. $\overline{R^2}$ increases when adding new regressor(or a group of regressors) if and only if the t value (F) for the individual regression(group of regressors) is more than 1.

6.6 Causal Mechanism

6.7 Confidence Interval for Prediction