Midterm Preperation

LIU Chenyuan

Spring 2024

- Midterm exam: April 14th, 7:00pm - 9:00pm.
- Location
  1. 6C201 if your student ID is 2017080182 - 2022010900
  2. 6C202 if your student ID is 2022010901 - 2022011461
  3. 6C300 if your student ID is 2022011462 - 2024403080
- You can bring a double-sided A4 paper with formulas (print or handwritten). No calculator is needed.
- Range: Ch1 - Ch5, no Stata command will be tested, but we may ask you to interpret Stata outputs.
- Office hour before the Midterm: April 11th, Thursday 3-5 pm, Lihua Building B629

Outline

Interpretation of Regression Model

Estimation of Parameters

Properties of the OLS Estimators

Hypothesis Testing of Parameters

Three Interpretations of A Regression Model

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

1. Descriptive: conditional mean

$$E(y|x) = \beta_0 + \beta_1 x.$$

2. Causal: β_1 represents the causal effect of x on y
3. Forecasting

Causal Interpretation: $E(u|x) = 0$

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

- u: factors affecting y other than x
- β_1 represents causal effect of x on y if u does not change when x changes
- Fix x and take expectation:

$$E[y|x] = \beta_0 + \beta_1 x + E[u|x].$$

Take derivatives with regard to x:

$$\frac{\partial E[y|x]}{\partial x} = \beta_1 + \frac{\partial E[u|x]}{\partial x} = \beta_1$$

if $E(u|x) = 0$.

Multiple Regression Model: Causal Interpretation

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u,$$

and $E(u|x) = 0$.

- β_1 represents holding fixed other factors, the change in y when x_1 increases by one unit.
- Partialling out interpretation: β_1 measures the relationship between y and x_1 after $x_2, ..., x_k$ have been partialled out.

Other Interpretation Issues

- $log(y) \sim x$: β_1 is approximately the fractional change in y when x changes by 1 unit
- Goodness of fit: R^2 represents the fraction of variation in the dependent variable explained by the model.

Outline

Method 1: Sample Analogue

$E(u|x) = 0$ implies:

Population expectations	Sample analogue
$E(u) = 0$	$\frac{1}{N}\sum_{i=1}^{N} \hat{u}_i = 0$
$E(x_1 u) = 0$	$\frac{1}{N}\sum_{i=1}^{N} x_{i1}\hat{u}_i = 0$
...	...
$E(x_k u) = 0$	$\frac{1}{N}\sum_{i=1}^{N} x_{ik}\hat{u}_i = 0$

Method 2: Ordinary Least Squares

- Minimize the sum of the residual square.
- Fitted value: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + ... + \hat{\beta}_k x_{ik}$.
- Residual: $\hat{u}_i = y_i - \hat{y}_i$
- We choose βs to minimize the sum of residual squares:

$$\min_{\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k} \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - ... - \hat{\beta}_k x_{ik})^2.$$

Method 3: FWL Theorem

$\hat{\beta}_j$ can be obtained by

1. Regress x_{ij} on other independent variables (including the intercept), obtain the residual $\hat{r}_{ij}$.
2. (Can skip) regress y_i on other independent variables (including the intercept), obtain the residual $\hat{r}_{iy}$.
3. Regress y_i (or $\hat{r}_{iy}$) on $\hat{r}_{ij}$. The resulting slope coefficient is $\hat{\beta}_j$.

OLS Estimators

- For simple linear regression model:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

- For multiple linear regression model:

$$\hat{\beta}_j = \frac{\sum_{i=1}^{N} \hat{r}_{ij} y_i}{\sum_{i=1}^{N} \hat{r}_{ij}^2}, \quad \forall j = 1, 2, ..., k.$$

Outline

Normal Equations

The following conditions hold for any sample of data and are true regardless of interpretations:

$$\sum_{i=1}^{N} \hat{u}_i = 0.$$

$$\sum_{i=1}^{N} \hat{u}_i x_{ij} = 0, \forall j = 1, 2, ..., k.$$

$$\sum_{i=1}^{N} \hat{u}_i \hat{y}_i = 0.$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + ... + \hat{\beta}_k \bar{x}_k.$$

Classical Linear Model Assumptions

Key: understand the meaning of each assumption.

1. MLR.1 Linear in Parameters
2. MLR.2 Random Sampling
3. MLR.3 No Perfect Collinearity (SLR. 3 Sample Variation in the Explanatory Variable)
4. MLR.4 Zero Conditional Mean: $E(u|x) = 0$
5. MLR.5 Homoskedasticity: $Var(u|x) = \sigma^2$
6. MLR.6 Normality: $u \sim N(0, \sigma^2)$

Gauss-Markov Theorem

Key: understand the meaning of each property; knowing how we prove them (unless not required)

- Under MLR.1-MLR.4, the OLS estimators are unbiased and consistent.
- Under MLR.1-MLR.5, the OLS estimators are BLUE. We can derive the variance formula and the asymptotic distribution of the OLS estimators.
  - Note the difference between standard error and standard deviation of $\hat{\beta}_j$.
- Under MLR.1-MLR.6, the OLS estimators are BUE. We can derive the sampling distribution of the OLS estimators.

Extension

- Underspecifying the model (excluding a relevant variable): affect unbiasedness — omitted variable bias
- Overspecifying the model (including irrelevant variables in a regression model): affect the variance
- Multicollinearity: high but not perfect correlation among parameters; affect the variance

Proof Hints

- Law of iterated expectations: $E(X) = E(E(X|Y))$.
- When proving something new, try to convert the problem to an existing problem which we have an answer.
- Be careful about the notation: expectation or sample mean? $\hat{u}_i$ or u_i?
- Feel free to use the results we already showed in class, for example, the FWL theorem, the Gauss-Markov Theorem etc., unless the question asked otherwise.

Outline

Interpretation of Regression Model

Estimation of Parameters

Properties of the OLS Estimators

Hypothesis Testing of Parameters

Testing Hypothesis About One Parameter

$$\begin{aligned} H_0: &\quad \beta_j = 0. \\ H_1: &\quad \beta_j \neq 0. \end{aligned}$$

Reject H_0 if

1. $|t_{\hat{\beta}_j}| > c$, where $t_{\hat{\beta}_j} = \hat{\beta}_j / se(\hat{\beta}_j)$
2. $p < \alpha$, where $p = 2P(T > |t|)$
3. 0 is not inside the $(1 - \alpha)$ confidence interval.

Testing Single Hypothesis About Multiple Variables

$$
\begin{aligned}
H_0: &\quad \beta_j = \beta_k. \\
H_1: &\quad \beta_j \neq \beta_k.
\end{aligned}
$$

- We can transform this question into a test of a single parameter for another regression model.
- Define $\theta = \beta_j - \beta_k$, then plug in the original regression equation, such that θ is a parameter. Test $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ in the new model.

Testing Multiple Linear Restrictions

$$\begin{aligned} H_0 : &\quad \beta_j = 0 \text{ and } \beta_k = 0. \\ H_1 : &\quad H_0 \text{ is not true} \end{aligned}$$

- Estimate the original model, obtain SSR_{ur}.
- Estimate the restricted model (the model when H_0 is true), obtain SSR_r.
- Calculate the F stat: $F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(N-k-1)}$. Reject H_0 if $F > c$ (or if $p < \alpha$.)

After the Midterm...

After the midterm: extensions from the basic model

1. Ch6: functional forms and prediction analysis
2. Ch7: binary independent variable
3. Ch8: heteroskedasticity
4. Ch12: serial correlation
5. Ch9: (mis)specification and measurement error
6. Ch15: instrumental variable
7. Ch17: limited dependent variable

Good Luck! ^_^

THE STUDENT VIEW:
AHH! I HAVE TO TAKE AN EXAM TOMORROW!
JORGE CHAM © 2018
THE PROFESSOR VIEW:
AHH! I HAVE TO GIVE AN EXAM TOMORROW!
WWW.PHDCOMICS.COM