**Chapter 1. Simple Regression Model**

$$y_t = \beta_1 + \beta_2 x_t + e_t, \quad t = 1, 2, \ldots, T.$$

**Sources of the error term** $e$

- the randomness in human behavior.
- effect of a large number of variables that have been omitted.
- measurement error in $y$.

**Assumptions about the model:**

1. $e_t$ is a random variable with $\mathrm{E}(e_t) = 0$.
2. $\mathrm{Var}(e_t) = \mathrm{E}(e_t^2) = \sigma^2$ (homoscedasticity).
3. $e_t$ and $e_s$ are independent for $t \neq s$. $\mathrm{Cov}(e_t, e_s) = 0$ for $t \neq s$.
4. $e_t$ and $x_t$ are independent. If $x_t$ is non-random, this assumption obviously holds.
5. Not all the values of $x_t$ are the same. At least one of the $x_t$ is different from others. This assumption ensures that there is some variation in $x$, i.e., $\sum_t (x_t - \bar{x})^2 \neq 0$.
6. (Normality) $e_t \sim N(0, \sigma^2)$.

# 1 Ordinary Least Squares

## 1.1 Estimation

**1 Ordinary Least Squares 1.1.1 Estimation**

- $e_t$ is the distance between the model (the regression line) and the observation.
- $$\min \sum_{t=1}^{T} e_t^2.$$
- We choose $\beta_1$ and $\beta_2$ so that
$$S(\beta_1, \beta_2) = \sum_{t=1}^{T} (y_t - \beta_1 - \beta_2 x_t)^2$$
is minimized.

## 1 Ordinary Least Squares

- How do you find the minimal for $x^2$? How about $(x-2)^2$?
- Slope $= 0$. (Should check the second-order condition. But will ignore in this class.)
- $\frac{d}{dw}(aw+b)^2 = 2a(aw+b)$.

Using "partial differentiation", one gets the partial derivatives of $S$ with respect to $\beta_1$ and $\beta_2$

$$\frac{\partial S}{\partial \beta_1} = 2\sum_{t=1}^{T}(y_t - \beta_1 - \beta_2 x_t)(-1) \tag{1}$$

$$\frac{\partial S}{\partial \beta_2} = 2\sum_{t=1}^{T}(y_t - \beta_1 - \beta_2 x_t)(-x_t) \tag{2}$$

- Set the above two equations to zero, and replace $\beta_1$ and $\beta_2$ by $b_1$ and $b_2$, we obtain

$$2\sum_{t=1}^{T}(y_t - b_1 - b_2 x_t)(-1) = 0 \tag{3}$$

$$2\sum_{t=1}^{T}(y_t - b_1 - b_2 x_t)(-x_t) = 0 \tag{4}$$

We call the above two equations the "normal equations."

- Note the above two equations are equivalent to

$$\sum_{t=1}^{T}\hat{e}_t = 0 \tag{5}$$

$$\sum_{t=1}^{T}x_t\hat{e}_t = 0 \tag{6}$$

- $$b_1 = \frac{1}{T}\sum_t y_t - \frac{1}{T}(\sum_t x_t)b_2 \equiv \bar{y} - b_2\bar{x}, \tag{7}$$

- $$b_2 = \frac{\sum_t (x_t - \bar{x})y_t}{\sum_t (x_t - \bar{x})^2}. \tag{8}$$

- In some textbooks you may find

$$b_2 = \frac{\sum_t (x_t - \bar{x})(y_t - \bar{y})}{\sum_t (x_t - \bar{x})^2} \tag{9}$$

or

$$b_2 = \frac{T\sum x_t y_t - (\sum x_t)(\sum y_t)}{\left[T(\sum x_t^2) - (\sum x_t)^2\right]}. \tag{10}$$

They are all equivalent.

**Remarks**

- Let the best fitted line be $\hat{y}_t = b_1 + b_2 x_t$.
- The least squares residual is defined as $\hat{e}_t = y_t - \hat{y}_t = y_t - b_1 - b_2 x_t$.
- Least squares *estimators* are general formulas and are random variables.
- Least squares *estimates* are numbers that are the observed values of random variables.
- Suppose that in the food consumption example we obtain the least squares estimates using a sample of 40 households and have $b_2 = 0.1283, b_1 = 40.7676$.
- The estimated (or fitted) regression line (Sample Regression Function) is

$$\hat{y}_t = 40.7676 + 0.1283 x_t. \tag{11}$$

## 1.2 Interpreting the Estimates

**1.1.2 Interpreting the Estimates**
$b_2$ and $b_1$.

## 1.3 Elasticities

**1.1.3 Elasticities**

- The income elasticity of demand is defined as

$$\eta = \frac{\text{percentage change in y}}{\text{percentage change in x}} = \frac{\Delta y / y}{\Delta x / x} = (\frac{\Delta y}{\Delta x})(\frac{x}{y})$$

- Ignoring the error term $e$, we get $\Delta y = \beta_2 \Delta x$. Hence,

$$\eta = \beta_2 (\frac{x}{y}) \tag{12}$$

- Note that the elasticity $\eta$ given above depends on $x$ and $y$. In practice economists often use $x = \bar{x}$ and $y = \bar{y}$.
- Thus, the elasticity at the mean values of $x$ and $y$ is

$$\hat{\eta} = b_2 (\frac{\bar{x}}{\bar{y}}) = 0.1283 \times \frac{698.00}{130.31} = 0.687$$

- The interpretation of this result is that, for 1% change in weekly household income will lead, on average, to approximately a 0.7% increase in weekly household expenditure on food. It suggests that food is a “necessity” rather than a “luxury”.

# 2 Prediction

**2 Prediction**

- Predict $y_0 = \beta_1 + \beta_2 x_0 + e_0$? It is essentially to estimate $y_0$.
- A natural candidate is $\hat{y}_0 = b_1 + b_2 x_0$ since $E(e_0) = 0$.

- Define the forecast error $f = y_0 - \hat{y}_0$. $E(f) = 0$.
- Example: $x_0 = 750$, $y_0 =$?
- $$\hat{y}_0 = b_1 + b_2 x_0 = 40.7676 + 0.1283(750) = \$136.98$$
- How about if $x_t =$ one of the data, what is the prediction? We call $\hat{y}_t$ the predicted value or the fitted value, $\hat{e}_t$ the residual.

# 3 Other Economic Models and Functional Forms

### 3 Functional Forms

- Linear regression models are "linear" in the parameters $\beta_1$ and $\beta_2$, not necessarily in the variables.
- Are the following models "linear" models?

$$\ln(y_t) = \beta_1 + \beta_2 \ln(x_t) + e_t$$

$$y_t = \beta_1 + \beta_2 x_t^2 + e_t$$

$$y_t = \beta_1 + \beta_2 (1/x_t) + e_t$$

If we use a log-log model: $\ln(y_t) = \beta_1 + \beta_2 \ln(x_t) + e_t$, then the elasticity

$$\eta = \frac{dy}{dx} \cdot \frac{x}{y} = \frac{d \ln y}{d \ln x} = \beta_2$$

Note that the elasticity of a log-log model is a constant (independent of $(x, y)$ values).

### Functional Forms

| Type | Nonlinear Form | Statistical Model | Marginal Effect ($\frac{dy}{dx}$) | Elasticity ($\frac{d\ln y}{d\ln x}$) |
|---|---|---|---|---|
| Linear | | $y_i = \beta_1 + x_i \beta_2 + e_i$ | $\beta_2$ | $\beta_2 \frac{x_i}{y_i}$ |
| Reciprocal | | $y_i = \beta_1 + \frac{1}{x_i} \beta_2 + e_i$ | $-\beta_2 \frac{1}{x_i^2}$ | $-\beta_2 \frac{1}{x_i y_i}$ |
| Log-Log | $y_i = \exp\{\beta_1\} x_i^{\beta_2} \exp\{e_i\}$ | $\ln y_i = \beta_1 + \beta_2 \ln x_i + e_i$ | $\beta_2 \frac{y_i}{x_i}$ | $\beta_2$ |
| Log-Linear (Exponential) | $y_i = \exp\{\beta_1 + x_i \beta_2 + e_i\}$ | $\ln y_i = \beta_1 + x_i \beta_2 + e_i$ | $\beta_2 y_i$ | $\beta_2 x_i$ |
| Linear-Log (Semilog) | $\exp\{y_i\} = \exp\{\beta_1 + e_i\} x_i^{\beta_2}$ | $y_i = \beta_1 + \beta_2 \ln x_i + e_i$ | $\beta_2 \frac{1}{x_i}$ | $\beta_2 \frac{1}{y_i}$ |
| Log-inverse | $y_i = \exp\left\{\beta_1 - \frac{1}{x_i} \beta_2 + e_i\right\}$ | $\ln y_i = \beta_1 - \frac{1}{x_i} \beta_2 + e_i$ | $\beta_2 \frac{y_i}{x_i^2}$ | $\beta_2 \frac{1}{x_i}$ |

Table 1: Summary of Some Conventional Functional Forms (P. 260 in GHJ1993)

### How to choose a functional form?

1. Economic theory.
2. What function form has been established in the literature?
3. Statistical tests.

# 4 Properties of the Least Squares Estimators

## 4.1 The least squares estimators as random variables

**4 Properties of the Least Squares Estimators 4.4.1 The least squares estimators as random variables**

- $b_1$ and $b_2$ are random variables.
- What are the means, variances and covariance of $b_1$ and $b_2$, and their distributions?
- Is there any "better" estimators than the least squares estimators $b_1$ and $b_2$?

## 4.2 The expected values of $b_1$ and $b_2$

**4.4.2 E($b_1$) and E($b_2$)**

- From equation (8), one can easily show that

$$\begin{aligned} b_2 &= \sum_t w_t y_t \\ &= \frac{\sum_t (x_t - \bar{x})(\beta_1 + \beta_2(x_t - \bar{x}) + \beta_2 \bar{x} + e_t)}{\sum_t (x_t - \bar{x})^2} \\ &= \beta_2 + \sum_t w_t e_t \end{aligned} \tag{13}$$

  where $w_t = (x_t - \bar{x}) / \sum_i (x_i - \bar{x})^2$.

- Note that $w_t$ is non-random (since $x_t$ is non-random in A4), we have

$$E(b_2) = E(\beta_2) + \sum_t E(w_t e_t) = \beta_2 + \sum_t w_t E(e_t) = \beta_2 \tag{14}$$

  since $E(e_t) = 0$.

- We used Assumptions 1, 4, and 5 in the above proof.
- $E(b_2) = \beta_2$, or in other words, $b_2$ is an unbiased estimator of $\beta_2$.
- $b_2 = \sum_t w_t y_t$, which is a linear combination of $y_t$'s. Hence $b_2$ is a linear estimator (linear in the dependent variable).

$$\begin{aligned} b_1 &= \bar{y} - b_2 \bar{x} = \beta_1 + \beta_2 \bar{x} + \bar{e} - b_2 \bar{x} \\ &= \beta_1 + (\beta_2 - b_2)\bar{x} + \bar{e}. \end{aligned}$$

So

$$E(b_1) = \beta_1.$$

## 4.3 The variances of $b_1$ and $b_2$

**4.4.3Var($b_1$) and Var($b_2$)**

- Based on equation (13):

$$\begin{aligned}\text{Var}(b_2) &= \text{Var}(\beta_2 + \sum_t w_t e_t) = \text{Var}(\sum_t w_t e_t) \\ &= \sum_t w_t^2 \text{Var}(e_t) = \sum_t w_t^2 \sigma^2 = \sigma^2 \sum_t w_t^2 = \frac{\sigma^2}{\sum_t (x_t - \bar{x})^2} \end{aligned} \tag{15}$$

  where in the last step we used

$$\sum_t w_t^2 = \sum_t [\frac{(x_t - \bar{x})^2}{\{\sum_i (x_i - \bar{x})^2\}^2}] = \frac{\sum_t (x_t - \bar{x})^2}{\{\sum_i (x_i - \bar{x})^2\}^2} = \frac{1}{\sum_t (x_t - \bar{x})^2}.$$

- Assumptions 2 and 3, in addition to 1, 4, and 5, are used here.

Similarly, one can show that

$$\text{Var}(b_1) = \sigma^2 [\frac{\sum_t x_t^2}{T \sum_t (x_t - \bar{x})^2}] \quad \text{and} \quad \text{Cov}(b_1, b_2) = \sigma^2 [\frac{-\bar{x}}{\sum_t (x_t - \bar{x})^2}]$$

**The Gauss-Markov Theorem**

Under assumptions of the linear regression model, the estimators $b_1$ and $b_2$ have the smallest variance among all linear unbiased estimators of $\beta_1$ and $\beta_2$. They are the **Best Linear Unbiased Estimators (BLUE)** of $\beta_1$ and $\beta_2$.

## 4.4 The probability distributions of the least squares estimators

**4.4.4 The probability distributions of the least squares estimators**

- Up to now we have not used the normality assumption yet.
- If we assume that $e_t$'s are normal random variables, then so is $b_2$ ($b_1$ as well) since $b_2 = \beta_2 + \sum_t w_t e_t$ is a linear combination of $e_t$'s.
- Under the normality assumption A6, we have

$$\begin{aligned} b_1 &\sim N(\beta_1, \frac{\sigma^2 \sum_t x_t^2}{T \sum_t (x_t - \bar{x})^2}) \\ b_2 &\sim N(\beta_2, \frac{\sigma^2}{\sum_t (x_t - \bar{x})^2}) \end{aligned} \tag{16}$$

- Without the normality assumption, we do not know the distribution of $b_1$ and $b_2$ (but we do know their means and variances).
- However, if the sample size is large ($T \to \infty$), the least squares estimators $b_1$ and $b_2$ still have a distribution that is very close to a normal distribution (by the Central Limit Theorem).

## 4.5 Estimating the variance of the error term

**4.4.5 Estimate $\sigma^2$**

- We assume we know $\sigma^2$ in $\text{Var}(b_i)$. But we don't.
- $\sigma^2 = \text{Var}(e_t) = \text{E}[e_t - E(e_t)]^2 = \text{E}(e_t^2)$ because $\text{E}(e_t) = 0$.
- Since $\text{E}(e_t^2)$ is the mean (average) value of $e_t^2$, it seems natural to use

$$\tilde{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2$$

  as an estimator for $\sigma^2$.

- But we don't observe $e_t$. Only $\hat{e}_t$ is estimated.
- $$\hat{\sigma}^2 = \frac{\sum_t \hat{e}_t^2}{T-2} \tag{17}$$
- It can be shown that $\text{E}(\hat{\sigma}^2) = \sigma^2$.

$$\begin{aligned}
\widehat{\text{Var}}(b_1) &= \hat{\sigma}^2 [\frac{\sum_t x_t^2}{T\sum_t (x_t - \bar{x})^2}] \\
\widehat{\text{Var}}(b_2) &= \frac{\hat{\sigma}^2}{\sum_t (x_t - \bar{x})^2} \\
\widehat{\text{Cov}}(b_1, b_2) &= \hat{\sigma}^2 [\frac{-\bar{x}}{\sum_t (x_t - \bar{x})^2}]
\end{aligned} \tag{18}$$

We call $\hat{\sigma}^2$ the MSE of the regression and $\hat{\sigma}$ the root MSE of the regression.

## 4.6 The least squares predictor

**4.4.6 The least squares predictor**

- Predict
$$y_0 = \beta_1 + \beta_2 x_0 + e_0 \tag{19}$$
- $$\hat{y}_0 = b_1 + b_2 x_0 \tag{20}$$
- The forecast error is

$$f \equiv \hat{y}_0 - y_0 = b_1 + b_2 x_0 - (\beta_1 + \beta_2 x_0 + e_0) = (b_1 - \beta_1) + (b_2 - \beta_2) x_0 - e_0 \tag{21}$$

- It is easy to show that $\text{E}(f) = E(\hat{y}_0 - y_0) = \text{E}(b_1 - \beta_1) + \text{E}(b_2 - \beta_2) - \text{E}(e_0) = 0 + 0 - 0 = 0$. Hence, $\hat{y}_0$ is an unbiased estimator for $y_0$. I.e., $\text{E}(\hat{y}_0) = y_0$.

$$\text{Var}(f) = \text{Var}(\hat{y}_0 - y_0) = \sigma^2[1 + \frac{1}{T} + \frac{(x_0 - \bar{x})^2}{\sum_t (x_t - \bar{x})^2}]$$

$$\widehat{\text{Var}}(f) = \widehat{\text{Var}}(\hat{y}_0 - y_0) = \hat{\sigma}^2[1 + \frac{1}{T} + \frac{(x_0 - \bar{x})^2}{\sum_t (x_t - \bar{x})^2}] \tag{22}$$

where the first term comes from the predictive error in $e_0$ and the second and the third terms come from the variances of $b_1$ and $b_2$ and their covariance.

The standard deviation of $f = \hat{y}_0 - y_0$ is

$$se(f) = \sqrt{\widehat{\text{Var}}(f)}$$

**Prediction in the food expenditure example**

We predicted the weekly expenditure on food for a household with weekly income $x_0 = \$750$. The predicted value was

$$\hat{y}_0 = b_1 + b_2 x_0 = 40.7676 + .1283(750) = \$136.98$$

Using our estimate $\hat{\sigma}^2 = 1429.2456$, the estimated variance of the forecast error is

$$\widehat{var}(f) = \hat{\sigma}^2[1 + \frac{1}{T} + \frac{(x_0 - \bar{x})}{\sum_t (x_t - \bar{x})^2}] = 1429.2456[1 + \frac{1}{40} + \frac{(750 - 698)^2}{1532463}] = 1467.4986$$

and

$$se(f) = \sqrt{\widehat{var}(f)} = \sqrt{1467.4986} = 38.3079.$$

# 5 Inference in the Simple Regression Model

## 5.1 Interval Estimation

**5 Inference in the Simple Regression Model 5.5.1 Interval Estimation**

- $$t = \frac{b_k - \beta_k}{se(b_k)} \sim t_{(T-2)}, \quad k = 1, 2. \tag{23}$$
- Using $Prob(|t| < t_{\alpha/2}) = 1 - \alpha$, we have the interval estimator

$$\text{Prob}\left[b_k - t_{\alpha/2} se(b_k) \le \beta_k \le b_k + t_{\alpha/2} se(b_k)\right] = 1 - \alpha. \tag{24}$$

## 5.2 Prediction intervals

**5 Inference in the Simple Regression Model 5.5.2 Prediction Interval Estimation**

- A $(1-\alpha) \times 100\%$ confidence interval, or prediction interval, for $y_0$ is

$$\hat{y}_0 \pm t_c se(f)$$

## 5.3 Hypothesis Testing

**5.5.3 Hypothesis Testing**

Components of Hypothesis Tests

1. Set up a null hypothesis that you want to test, $H_0$.
2. Set up an alternative hypothesis, $H_1$.
3. Choosing a test statistic.
4. Find out the rejection range.
5. State your conclusion.

- Under $H_0$,

$$t = \frac{b_2 - \beta_2}{se(b_2)} = \frac{b_2}{se(b_2)} \sim t_{(T-2)}$$

  i.e., $t$ is centered at (close to) zero under $H_0$.

- Under $H_1^a$, $b_2 \neq \beta_2$, and $t$ does not have a $t_{(T-2)}$ distribution.
- Rejection rule: for a given data set, we compute $t$, if this $t$ value does not look like a random draw from a $t_{(T-2)}$ distribution, we reject $H_0$. We do not reject $H_0$ otherwise.
- We need to choose a significant level $\alpha$ (type I error). Usually $\alpha = .01$, 0.05 or 0.10.

**The food expenditure example**

For testing $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$. By selecting $\alpha = 0.05$, we get the critical value $t_c = 2.021$.

For the food expenditure data, we get $t = b_2/se(b_2) = .1283/.0305 = 4.21$.

**The p-value of a hypothesis test**

When reporting the outcome of statistical hypothesis tests it has become common practice to report the p-value of a test.

For the above example, the t-value is 4.20. $P[|t_{(38)}| > 4.20] = .000155$. Hence, as long as we select $\alpha > .000155$, we will reject $H_0$ and in favor of $H_1^a$.

**One-sided tests**

Read pp. 139-140 in GHJ1993.

**Comments on stating null and alternative hypotheses**

- If we fail to reject $H_0 : \beta_2 = 0$, we probably would fail to reject $H_0 : \beta_2 = 0.01$ or $H_0 : \beta_2 = 0.02$ as well. That's why we cannot say we *accept* $H_0$, because you can't accept conflicting statements.
- Rejecting a null hypothesis is a stronger conclusion than failing to reject it.
- So the null hypothesis is usually stated in such a way that if our theory is correct then we will reject the null hypothesis.
- For example, if economic theory suggests that there should be a positive relationship between income and food expenditure, we would set up $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 > 0$.

# 6 Reporting Results

**6 Reporting Results**

- Recall that $\hat{e}_t = y_t - \hat{y}_t$. Therefore
$$y_t = \hat{y}_t + \hat{e}_t$$
- Subtracting the sample mean $\bar{y}$ from both sides of the above equation we get
$$y_t - \bar{y} = (\hat{y}_t - \bar{y}) + \hat{e}_t \tag{25}$$
-
$$\begin{aligned}\sum_t (y_t - \bar{y})^2 &= \sum_t [(\hat{y}_t - \bar{y}) + \hat{e}_t]^2 \\ &= \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t \hat{e}_t^2 + 2\sum_t (\hat{y}_t - \bar{y})\hat{e}_t \\ &= \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t \hat{e}_t^2 \qquad (26)\end{aligned}$$

$$\sum_t (y_t - \bar{y})^2 = \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t \hat{e}_t^2 \tag{27}$$
$$TSS = ESS + RSS \tag{28}$$
$$\sum_t (y_t - \bar{y})^2 = \text{total sum of squares } = TSS \tag{29}$$
$$\sum_t (\hat{y}_t - \bar{y})^2 = \text{explained sum of squares } = ESS \tag{30}$$
$$\sum_t \hat{e}_t^2 = \text{residual (unexplained) sum of squares } = RSS \tag{31}$$

- The **coefficient of determination** (or goodness-of-fit) $R^2$ is defined by
$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \tag{32}$$
- $0 \leq R^2 \leq 1$ if there is an intercept in the regression.
- The closer $R^2$ is to 1, the better our linear model fits the data. If $R^2 = 1$, then all the sample data fall exactly of the fitted straight line.

The correlation between two random variables $X$ and $Y$ is
$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

For a given sample data pairs $(x_t, y_t)$, $t = 1, ..., T$. The sample correlation coefficient $r^2$ is defined as (a sample analogue of $\rho$)
$$r^2 = \frac{(\sum_t (x_t - \bar{x})(y_t - \bar{y}))^2}{[\sum_t (x_t - \bar{x})^2][\sum_t (y_t - \bar{y})^2]} \tag{33}$$

It can be shown that $r^2 = R^2$ for the simple regression model.

## 6.1 Summarizing regression results

### 6.6.1 Summarizing regression results

- Usually we report the results in a table or in equation format.
-
$$\begin{aligned}\widehat{FOOD} &= 40.77 + 0.128 INCOME \\ & \quad (1.84) \quad (4.28) \\ & \quad T = 40, R^2 = 0.317,\end{aligned}$$

  where the numbers in the parentheses are the $t$ statistics.
-
$$\begin{aligned}\widehat{FOOD} &= 40.77 + 0.128 INCOME \\ & \quad (22.1) \quad (0.031) \\ & \quad T = 40, R^2 = 0.317,\end{aligned}$$

  where the numbers in the parentheses are standard errors.

```
.reg food income
      Source |       SS       df       MS              Number of obs =      40
-------------+------------------------------           F(  1,    38) =   17.65
       Model |   25221.219     1   25221.219           Prob > F      =  0.0002
    Residual |  54311.3331    38  1429.24561           R-squared     =  0.3171
-------------+------------------------------           Adj R-squared =  0.2991
       Total |  79532.5521    39  2039.29621           Root MSE      =  37.805
------------------------------------------------------------------------------
        food |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .1282886   .0305393     4.20   0.000     .0664651    .1901121
       _cons |   40.76756   22.13865     1.84   0.073    -4.049799    85.58493
------------------------------------------------------------------------------
```

- $TSS = ESS + RSS$.
- $Root\ MSE = \sqrt{RSS/d.f.}$.
- $R^2$.
- Adjusted $R^2$ (leave for multiple regression).
- Coefficient, Standard Error, $t$, P of $t$.
- Confidence interval.

# 7 The Basics of the Asymptotic Theories: An Example

**7 The Basics of the Asymptotic Theories: An Example**

Consider the following regression model: $y_i = \gamma z_i + e_i$ $(i = 1, \ldots, n)$ where all the quantities are scalars. We maintain the following assumptions: $e_i \sim i.i.d.(0, \sigma^2)$; $z_i$ and $e_i$ are uncorrelated; $z_i$'s are independent of each other; plim $\frac{\sum_{i=1}^n z_i^2}{n} = \sigma_z^2$, which is finite. The OLS estimator is given by $\hat{\gamma} = \frac{\sum z_i y_i}{\sum z_i^2}$.

1. Show that $\hat{\gamma}$ is a consistent estimator of $\gamma$.
2. Show that $\hat{\gamma}$ is asymptotically normally distributed.

Proof:

1. Note that we can rewrite $\hat{\gamma} = \frac{\sum z_i y_i}{\sum z_i^2} = \gamma + \frac{\sum z_i e_i}{\sum z_i^2}$.

   Let's focus on the second term on the RHS for a minute.

   $\mathrm{E}(\frac{\sum z_i e_i}{\sum z_i^2}) = 0$.

   $\mathrm{Var}(\frac{\sum z_i e_i}{\sum z_i^2}) = \frac{\sigma^2}{\sum z_i^2} = \frac{1}{n}\frac{\sigma^2}{\sum z_i^2/n} \to \frac{1}{n}\frac{\sigma^2}{\sigma_z^2} \to 0$ when $n \to \infty$.

   So, by the Mean Square Theorem, plim $\frac{\sum z_i e_i}{\sum z_i^2} = 0$, which implies that plim $\hat{\gamma} = \gamma + 0 = \gamma$.

   So $\hat{\gamma}$ is a consistent estimator.

2. Define $w_i = e_i z_i$. So $\hat{\gamma} - \gamma = \frac{1}{\sum z_j^2/n}\frac{1}{n}\sum w_i = \frac{1}{\sum z_j^2/n}\bar{w}$. Note that plim $\frac{1}{\sum z_j^2/n} = \frac{1}{\sigma_z^2}$.

   $\mathrm{E}(w_i) = 0$. $\mathrm{Var}(w_i) = \sigma^2 z_i^2$. So by the CLT, $\sqrt{n}\bar{w} \xrightarrow{d} N(0, \sigma^2\sigma_z^2)$.

   Thus, $\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, \frac{\sigma^2}{\sigma_z^2})$.

# 8 The Method of Moments

**8 The Method of Moments**

- Principle: match the population moments (mean, variance, covariance, etc.) with their sample counterparts.
- Recall that we have two assumptions about the regression model: $\mathrm{E}(e_t) = 0$ and $\mathrm{Cov}(x_t, e_t) = 0$ (from their independence). The sample counterpart of $e_t$ is

  $$\hat{e}_t = y_t - b_1 - x_t b_2.$$

  So the sample counterparts of the two population assumptions are

  $$\frac{1}{T}\sum \hat{e}_t = 0 \quad (34)$$
  $$\frac{1}{T}\sum \hat{e}_t x_t = 0 \quad (35)$$

- Compare these two equations with the two normal equations for OLS: they are identical! So the MM estimators are the same as the OLS estimators.
- For some models they may be different.
- There are two unknowns and two equations (moment conditions) in the above example. What if we have more moment conditions than unknowns? Generalized Method of Moments (GMM) (using a weight matrix for different moment conditions).

The Method of Maximum Likelihood Estimation

# 9 Appendix A: Matrix Notations

**9 Appendix A: Matrix Notations**

- Read the Appendix to Chapter 2 in Maddala (2001) for basic matrix notations.
- Read the Appendix to Chapter 3 for the simple regression.
- A matrix is a rectangular array of elements. $A = \{a_{ij}\}_{m\times n}$.

- A vector is a one-row matrix or a one column matrix.
- $A+B$, $A-B$.
- Product between a scalar and a matrix: $cA$.
- Product between matrices: $AB$. Typically $AB \neq BA$. The order of matrices.
- Transpose (prime). $(ABC)' = C'B'A'$.
- Identity matrix $I_n$. $I_N$.
- Inverse: $A^{-1}A = AA^{-1} = I$.
- Determinant: $|A|$.