Eco 4306 Economic and Business Forecasting

Chapter 2: Review of the Linear Regression Model

Outline

Review

- conditional density and conditional moments
- linear regression model
- ordinary least squares method
- hypothesis testing

2.1 Conditional Density and Conditional Moments

- consider two random variables Y and X, e.g. annual consumption Y and annual income X, for which we would like to understand their relation
- suppose we want to answer the following questions
  1. For people with income \$60,000, what is the expected average consumption?
  2. For people with income \$80,000, what is the expected variability in consumption?
  3. For people with income \$40,000, what is the probability of spending at most \$30,000?
- for all three, we first need the **conditional probability density function** of consumption given income $f(Y|X)$ in order to
  1. calculate the **conditional expectation** of consumption given income

$$\mu_{Y|X=60\text{k}} = E(Y|X = 60\text{k}) = \int_{-\infty}^{\infty} Y f(Y|X = 60\text{k}) dY$$

  2. calculate the **conditional variance** of consumption given income

$$\sigma^2_{Y|X=80\text{k}} = var(Y|X = 80\text{k}) = \int_{-\infty}^{\infty} (Y - \mu_{Y|X=80\text{k}})^2 f(Y|X = 80\text{k}) dY$$

  3. calculate the **conditional probability** of consumption given income

$$P(Y \leq 30\text{k}|X = 40\text{k}) = \int_{-\infty}^{30\text{k}} f(Y|X = 40\text{k}) dY$$

2.1 Conditional Density and Conditional Moments

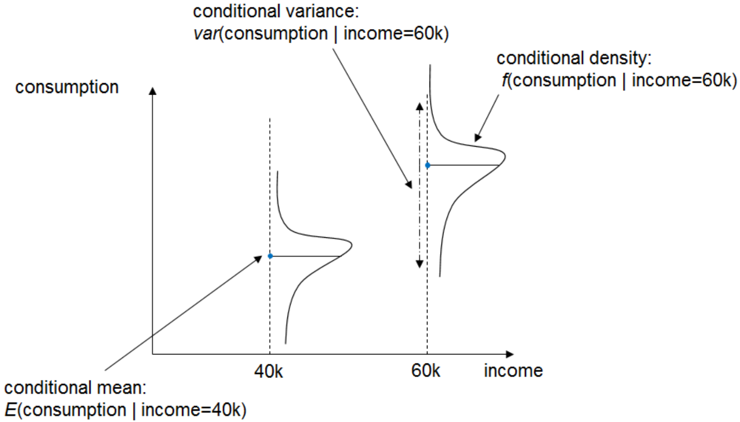
- sample with 50 households: 10 households in each of the 5 income brackets

income X, in thousands of \$	20	40	40	60	100
	19	25	44	55	55
	15	32	35	60	77
	11	37	55	75	88
	13	19	57	45	42
consumption Y	9	20	58	68	82
in thousands of \$	8	32	59	73	90
	18	31	42	71	77
	16	38	38	49	67
	12	18	33	56	60
	11	26	47	71	43
conditional mean of consumption $\mu_{Y\|X}$	13.20	27.80	46.80	62.30	68.10
conditional variance of consumption $\sigma^2_{Y\|X}$	13.73	53.29	98.18	114.90	308.54

2.1 Conditional Density and Conditional Moments

- conditional probability of Y and any conditional moment of Y are functions of X, their values change with changing value of X
- that is, we have $E(Y|X) = g_1(X)$, $var(Y|X) = g_2(X)$
- if Y and X are independent, X does not affect Y, thus conditional density is equal to marginal density, $f(Y|X) = f(Y)$, and conditional moments are equal to unconditional moments, $E(Y|X) = E(Y)$ and $var(Y|X) = var(Y)$
- if $E(Y|X) = E(Y)$ correlation between Y and X is zero, $\rho_{YX} = 0$.
  but: converse is not always true, because correlation refers only to linear dependence - it is possible that $\rho_{YX} = 0$ and the conditional mean may depend on X in a nonlinear fashion, for example, $E(Y|X) = a + bX^2$
- the main idea of forecasting is to choose the best functions that describe the conditional mean, the conditional variance, or any other conditional moment

2.1 Conditional Density and Conditional Moments



2.2 Linear Regression Model

- suppose we have data on consumption Y and income X on all individuals in some population
- suppose our goal is to answer questions about the average consumption for people with different levels of income
- that is, we are interested in the conditional mean of consumption given a level of income $E(Y|X)$

2.2 Linear Regression Model

simple **linear regression model**

- suppose that relation between Y and X in the population is given as

$$Y = \beta_0 + \beta_1 X + u$$

- terminology
  - Y is the dependent variable or response variable or regressand
  - X is the independent variable or explanatory variable or regressor
  - β_0 and β_1 are constant regression coefficients
  - u is the stochastic disturbance or error term
- interpretation of regression coefficients
  - dependence between Y and X is given by the coefficient β_1
  - derivative of Y with respect to X is regression coefficient β_1, i.e. $\frac{dY}{dX} = \beta_1$
  - marginal change in X causes a change in Y equal to β_1, i.e. $\Delta Y = \beta_1 \Delta X$

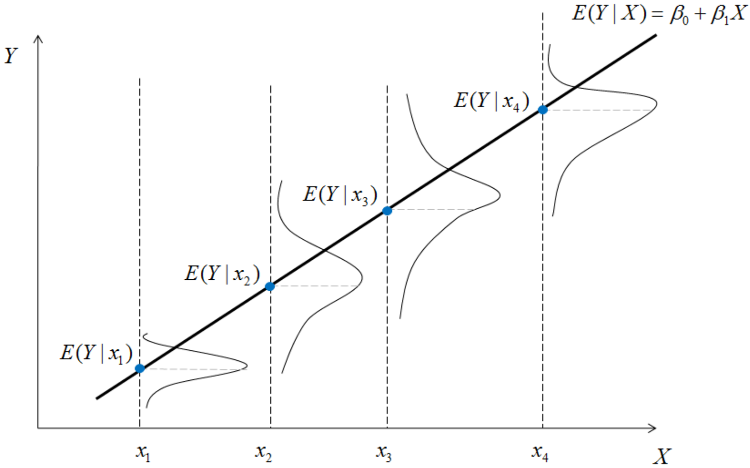
2.2 Linear Regression Model

- we observe Y and X, but the error term is unobservable
- error term accounts for any measurement errors in X and Y
- error term also captures the effect of variables that affect Y but are not explicitly accounted for in the model
- example: interest rates affect consumption, people defer present consumption in favor of future consumption when the interest rate on savings is high enough, and borrow less to spend if the interest rate on loans is high
- since we include an intercept β_0 in the model any omitted variables contained in the error term has *on average* no impact on Y thus, $E(u) = 0$
- important assumption: error term should not depend on the regressor X, that is, $E(u|X) = E(u)$, otherwise the estimate of β_1 becomes biased
- example: interest rates, implicitly contained in the error term u, are not correlated with different levels of income Y
- we then have

$$E(Y|X) = E(\beta_0 + \beta_1 X + u) = \beta_0 + \beta_1 X + E(u|X) = \beta_0 + \beta_1 X$$

2.2 Linear Regression Model

- $Y = \beta_0 + \beta_1 X + u$ and $E(Y|X) = \beta_0 + \beta_1 X$ since $E(u|X) = 0$



2.2 Linear Regression Model

multiple regression model

- relation between dependent variable Y and explanatory variables $X_1, X_2, \ldots, X_k$ is given by
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$
- under assumptions $E(u|X_1, X_2, \ldots, X_k) = E(u) = 0$ we get
$$E(Y|X_1, X_2, \ldots X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$
that is, conditional expectation of Y given fixed values for all k regressors is a linear function of the regressors

2.2 Linear Regression Model

interpretation of regression coefficients in $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$

- **in a multiple regression model β_j captures the partial marginal effect of regressor X_j on the dependent variable keeping the remaining regressors fixed**
- thus it is in the spirit of *ceteris paribus* common in comparative statics in macro and microeconomics

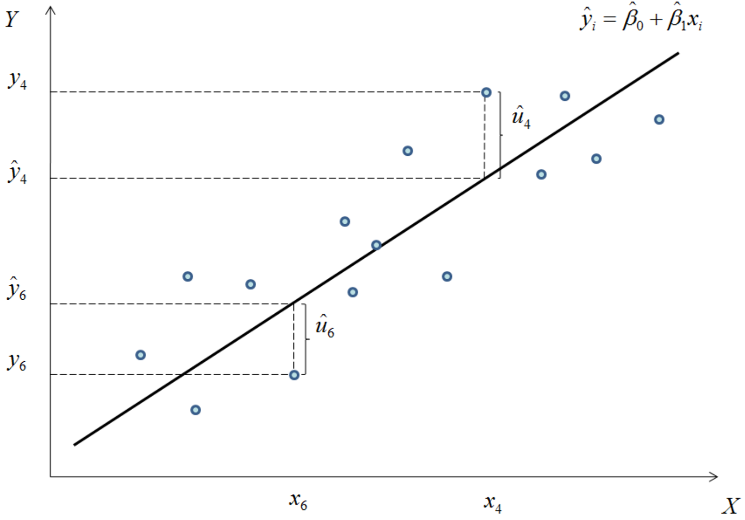
example

- suppose that $k = 3$ and that X_1 is income, X_2 net worth, and X_3 interest rate
- the effect of a marginal change in income on consumption, holding both net worth and interest rate fixed, is β_1
- in other words, β_1 is purely the effect of marginal changes in income, net of any change in other variable that may be highly correlated with X_1

2.3 Estimation: Ordinary Least Squares

- regression coefficients $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ in the linear regression model are unknown and need to be estimated
- one way how to estimate them: **ordinary least squares** (OLS) method
- to demonstrate the method consider simple linear regression model with one explanatory variable
$$y_i = \beta_0 + \beta_1 x_i + u$$
- suppose that we draw a **random sample** of n observations on Y and X that we denote by $(x_i, y_i),\ i = 1, 2, \ldots, n$
- example: x_i is income and y_i is consumption of household i, and we have this information on n households randomly chosen from population
- objective: find the “best” line for sample conditional mean $E(Y|X)$

2.3 Estimation: Ordinary Least Squares



2.3 Estimation: Ordinary Least Squares

- objective: find the "best" line for sample conditional mean $E(Y|X)$
- we will first need to define more rigorously what "best" means
  https://www.youtube.com/watch?v=WaaANll8h18

- suppose the estimates of true coefficients (β_0, β_1) are $(\hat{\beta}_0, \hat{\beta}_1)$ so that the **fitted value** for point i is $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- vertical distance between the actual observation y_i and the fitted value $\hat{y}_i$ is called the **residual** $\hat{u}_i$
- we would like to find a line so that the distance between the fitted values $\hat{y}_i$ and the data y_i is the smallest possible
- ordinary least squares method: choose $\hat{\beta}_0, \hat{\beta}_1$ that minimize the sum of squared residuals

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \hat{u}_i^2 = \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

- we choose the sum of square of the residuals so larger deviations lead to proportionally larger penalties
- note: there are other estimation methodologies with different optimality criteria, e.g. minimizing the sum of absolute residuals, maximum likelihood method, ...

2.3 Estimation: Ordinary Least Squares

- in a multiple regression model

$$\min_{\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_k} \sum_{i=1}^{n} \hat{u}_i^2 = \min_{\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_k} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \ldots - \hat{\beta}_k x_{ki})^2$$

- *sample regression line*

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_k x_{ki}$$

  is an estimate of *population regression line*

$$E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

- in population $\beta_0, \beta_1, \ldots, \beta_k$ are unknown constant coefficients (fixed parameters)
- in sample $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are estimators, random variables - given a different sample, OLS estimation we will yield different estimates $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$
- as with any random variable, we are interested in their moments, such as mean and variance, or/and their probability density functions

2.3.1 R^2 and Adjusted R^2

- random variable Y can be decomposed into both a systematic component and an unsystematic component, $Y = E(Y|X) + u$
- after we estimate the model we obtain the sample counterpart of the decomposition $Y = E(Y|X) + u$ using $y_i = \hat{y}_i + \hat{u}_i$
- **total sum of squares**

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

  is the sample variation in dependent variable y with respect to sample average $\bar{y}$

- **sum of squares explained by the model**

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

  is the sample variation in fitted values $\hat{y}$ with respect to the sample average $\bar{y}$

- **sum of squared residuals**

$$SSR = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \hat{u}_i^2$$

  is the sample variation in the residuals $\hat{u}$

2.3.1 R^2 and Adjusted R^2

- it can be proven that $SST = SSE + SSR$ so that the total variation can be decomposed into the explained variation (due to the model) and the unexplained variation (due to the residual)
- coefficient of determination R^2 provides a measure of goodness of fit - how well the model is on explaining the variability of Y

2.3.1 R^2 and Adjusted R^2

- the R-**squared** is defined as

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

- thus $0 \leq R^2 \leq 1$, and if the model is a good fit and fully explains the total variation of the dependent variable $SSE = SST$ and $R^2 = 1$; if the model is a very poor fit and does not at all explain the total variation $SSE = 0$ and $R^2 = 0$
- when new regressors are added R^2 will never decrease

2.3.1 R^2 and Adjusted R^2

- the **adjusted** R-**squared** provides a way to avoid adding irrelevant regressors

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$$

- if an irrelevant regressor is included in the model k will go up and $\bar{R}^2$ will go down, which will indicate that the new regressor is worthless
- introduction of degrees of freedom $n-k-1$ can be interpreted as a penalty function that balances the inclusion of more regressors against the quality of the information that they provide to explain the variability of the dependent variable

2.3.2 Linearity and OLS

- suppose we want to model consumption Y as a quadratic function of income X, so that

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u$$

- marginal propensity to consume then depends on level of income:

$$\frac{dY}{dX} = \beta_1 + 2\beta_2 X$$

  and if $\beta_1 > 0$, $\beta_2 < 0$ then higher level of income leads to lower marginal propensity to consume

- this model can be written as a linear regression model and be estimated by OLS even though Y is a nonlinear function of X
- we just need to define $W \equiv X^2$, so that

$$Y = \beta_0 + \beta_1 X + \beta_2 W + u$$

2.3.2 Linearity and OLS

- some other nonlinear specifications

  log-log
  $$\log Y = \beta_0 + \beta_1 \log X + u$$
  log-level
  $$\log Y = \beta_0 + \beta_1 X + u$$
  level-log
  $$Y = \beta_0 + \beta_1 \log X + u$$
  are also linear regression models, by redefining regressors and/or regressands
- but if
  $$Y = \frac{1}{1 + \beta_0 e^{\beta_1 X}}$$
  dependent variable Y is not a linear function of the regression coefficients β_0 and β_1 and the model cannot be estimated by OLS (**it's a horse of a different color**)

2.3.2 Linearity and OLS

- bottom line: **any specification falls into the realm of linear regression analysis as long as the dependent variable or some transformation of the dependent variable is a linear function of the regression coefficients**

2.3.2 Linearity and OLS

- if $Y = \beta_0 + \beta_1 X + u$ then
$$\frac{dY}{dX} = \beta_1$$
and Y increases by β_1 units if X increases by 1 unit
- if $\log Y = \beta_0 + \beta_1 \log X + u$, using calculus $d \log Y = dY/Y$ and we have
$$\frac{d \log Y}{d \log X} = \frac{dY/Y}{dX/X} = \beta_1$$
so Y increases by β_1 percent if X increases by 1 percent
e.g. if $\beta_1 = 3.7$ then Y increases by 3.7% when X increases by 1%
this is the **elasticity** concept prevalent in micro- and macroeconomics
- if $\log Y = \beta_0 + \beta_1 X + u$ then
$$\frac{d \log Y}{dX} = \frac{dY/Y}{dX} = \beta_1$$
and Y increases by $100 \times \beta_1$ percents if X increases by 1 unit
e.g. if $\beta_1 = 3.7$ then Y increases by 370% when X increases by 1 unit
- if $Y = \beta_0 + \beta_1 \log X + u$ then
$$\frac{dY}{d \log X} = \frac{dY}{dX/X} = \beta_1$$
and Y increases by $\beta_1/100$ units if X increases by 1 percent
e.g. if $\beta_1 = 3.7$ then Y increases by 0.037 units when X increases by 1%

2.3.3 Gauss-Markov Theorem

- if we repeatedly draw samples of (X, Y) from a given population and calculate the OLS estimates $\hat{\beta}$ for each sample, these estimates vary from sample to sample - they are subject to sample variation
- OLS estimates are realizations of a random variable that we call the OLS estimator
- in practice, given the nature of the non-experimental data in economics and business, we mostly work with only one sample
- the question is thus how precisely are true parameters β estimated
- in other words: how large is the variance $var(\hat{\beta})$, capturing our uncertainty regarding the true unknown value β?
- the answer is given by Gauss-Markov Theorem

2.3.3 Gauss-Markov Theorem - Assumptions

A1 Linearity: population regression model is linear in regression coefficients

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

A2 Zero Conditional Mean: conditional on regressors, expected value of error term is 0

$$E(u|X_1, X_2, \ldots, X_k) = 0$$

A3 Homoscedasticity: conditioning on regressors, variance of the error term is constant

$$var(u|X_1, X_2, \ldots, X_k) = \sigma_u^2$$

which also implies

$$var(Y|X_1, X_2, \ldots, X_k) = E[(Y - E(Y|X))^2|X] = E[u^2|X] = var(u|X_1, X_2, \ldots, X_k) = \sigma_u^2$$

A4 No Serial Correlation: in a regression model where data are gathered over time

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t$$

the errors should not be correlated over time and thus

$$cov(u_t, u_{t-l}) = 0 \qquad l = \pm 1, \pm 2, \ldots$$

A5 No Perfect Collinearity: There is not an exact linear relation among the regressors e.g. $X_2 = X_1/1000$ or $X_3 = X_2 + X_4$. With perfect collinearity, model cannot be estimated, one of the regressors needs to be removed.

A6 Sample Variation: no regressor can be constant for all observations in the sample

$$var(X_j) > 0 \qquad j = 1, 2, \ldots, k$$

2.3.3 Gauss-Markov Theorem

Gauss-Markov theorem: Under assumptions A1 through A6, the OLS estimators $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ are the **best linear unbiased estimators (BLUE)** of their respective population regression coefficients $(\beta_0, \beta_1, \ldots, \beta_k)$.

- linear - estimator $\hat{\beta}$ is a linear function of the dependent variable Y
- unbiased - expected value of the OLS estimator $\hat{\beta}$ is the corresponding population regression coefficient β

$$E(\hat{\beta}_j) = \beta_j \qquad j = 1, 2, \ldots, k$$

- best - variance of the OLS estimator $\hat{\beta}_j$ is the smallest among the linear and unbiased estimators obtained using other methods than OLS

$$var(\hat{\beta}_j) \leq var(\tilde{\beta}_j) \qquad j = 1, 2, \ldots, k$$

  this is also called **efficiency** property
- only A1, A2, A5, A6 are needed to guarantee that OLS is unbiased - even in the presence of heteroscedasticity and serial correlation, when A3 and A4 do not hold, the OLS estimator is still unbiased
- A3 and A4 are needed in addition to assumptions A1, A2, A5, A6 for OLS to be efficient (have lowest variance among linear unbiased estimator)

2.3.3 Gauss-Markov Theorem

- **Gauss-Markov Theorem**: OLS estimator $\hat{\beta}_{OLS}$ have lowest variance among linear unbiased estimator as long as errors are not heteroscedastic or serially correlated
- but in many instances economic data *are* heteroscedastic and serially correlated
- it is possible to modify the variance of the OLS estimator to account for heteroscedasticity and serial correlation
- in practice, we often use **Newey-West HAC standard errors** which are robust against heteroscedasticity and serial correlation
- efficiency (minimum variance) of the OLS estimator is lost when robust variances are used, but OLS estimator remains unbiased
- advanced estimation methods provide more efficient estimators than the OLS in the presence of heteroscedasticity and/or serial correlation - we will introduce **maximum likelihood** estimation later

2.3.4 An Example: House Prices and Interest Rates

- Question: Are house prices responsive to interest rates?
- Hypothesis: Low mortgage rates provide an incentive for consumers and investors to buy real estate. If the supply of houses grows slowly, so a strong demand for homes will put upward pressure on house prices. We thus expect lower mortgage rates to be associated with higher house prices.

2.3.4 An Example: House Prices and Interest Rates

- to investigate whether our hypothesis is correct we next perform regression analysis using data on house prices and mortgage interest rates
- data on house prices and mortgage interest rates: annual time series data from 1971 to 2007 for regional and national quarterly house price indexes and 30-year fixed rate on conventional mortgage loans, downloaded from Freddie Mac's website, http://www.freddiemac.com

2.3.4 An Example: House Prices and Interest Rates

- open EViews, create annual workfile, with start date 1971 and end date 2007
- import data from Microsoft Excel file **Table02_3_Data.xls** into EViews: click on **File**, select **Import**, then **Import from file**, and follow the instructions
- consider following multiple regression model

$$\Delta p_t = \beta_0 + \beta_1 \Delta p_{t-1} + \beta_2 \Delta r_{t-1} + u_t$$

  where Δp_t is the percentage change in the national house price index, and Δr_t is the percentage change in the 30-year fixed mortgage rate
- to estimate it in EViews first click on **Object**, select **New Object**, then **Equation**, in Specification screen enter **D_Price c D_Price(-1) D_Rate(-1)** and in Options screen under Coefficient covariance choose HAC (Newey-West)
- note: lagged change Δp_{t-1} is included since house prices tend to move slowly and changes from one period to the next are not abrupt - this persistence concept will be explained in detail later when we get to time series models

2.3.4 An Example: House Prices and Interest Rates

- since $\hat{\beta}_2 = -0.07$ there is an inverse relation between house price growth and changes in interest rates so our hypothesis seems to be validated
- however, we need to examine how statistically significant this finding is

Dependent Variable: D_Price
Method: Least Squares
Sample (adjusted): 1973 2007
Included observations: 35 after adjustments
Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.398569	0.541453	2.582990	0.0146
D_Price(-1)	0.747250	0.096000	7.783849	0.0000
D_Rate(-1)	-0.069321	0.045878	-1.510987	0.1406

R-squared	0.476213	Mean dependent var	6.255417
Adjusted R-squared	0.443476	S.D. dependent var	3.440050
S.E. of regression	2.566295	Akaike info criterion	4.804620
Sum squared resid	210.7479	Schwarz criterion	4.937936
Log likelihood	-81.08085	F-statistic	14.54676
Durbin-Watson stat	1.631411	Prob(F-statistic)	0.000032

$\hat{\beta}_0$
$\hat{\beta}_1$
$\hat{\beta}_2$

$\hat{\sigma}_{\hat{u}}$

SSR

$\hat{\sigma}_{\hat{\beta}}$ robust against heteroscedasticity and serial correlation

$\bar{y}$

$\hat{\sigma}_y$

2.4 Hypothesis Testing in a Regression Model

- having estimated a regression model, it is of interest to test the statistical significance of the regression coefficients
- for example, we have obtained that the marginal effect of interest rates on house prices is negative, with $\hat{\beta}_2 = -0.07$
- this $\hat{\beta}_2$ is only an estimate of true β given our relatively small sample with 35 observations
- we could thus ask whether more data would reveal that the true value β_2 is 0, so that on average, house prices do not react to changes in interest rates

2.4 Hypothesis Testing in a Regression Model

main idea for hypothesis testing

- we first form of a null hypothesis H_0 which is the claim that we would seek to disprove, and state an alternative hypothesis H_1 to accept when the null is rejected
- using tests we assess whether there is enough evidence in the data to reject the hypothesis H_0 or fail to reject H_0

we will review two tests

- the t-ratio test for single hypothesis
  for example $H_0 : \beta_1 = 0$
- the F-test for joint hypothesis
  for example $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

2.4.1 The t-ratio Test

- consider first the t-ratio test for single hypothesis
- we can test H_0 against **one-sided** or **two-sided** alternative hypothesis

  case (1): $H_0 : \beta_j = c$ against one-sided alternative $H_1 : \beta_j > c$
  case (2): $H_0 : \beta_j = c$ against one-sided alternative $H_1 : \beta_j < c$
  case (3): $H_0 : \beta_j = c$ against two-sided alternative $H_1 : \beta_j \neq c$

2.4.1 The t-ratio Test

- once we have chosen hypothesis to test, we need a test statistic and a decision rule
- to construct test statistic first note that standard deviation of the OLS estimator σ_{β_j} is a function of the error variance σ_u
- σ_u is not directly observable and needs to be estimated
- when we replace σ_u by its estimate $\hat{\sigma}_{\hat{u}}$ and thus σ_{β_j} becomes $\hat{\sigma}_{\beta_j}$ the pdf the estimator of the variances changes from Normal to Student-t with $n-k-1$ degrees of freedom
$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\beta_j}} \sim t_{n-k-1}$$
- we thus construct the t-ratio test statistic which will be Student-t distributed only if H_0 is true
$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - c}{\hat{\sigma}_{\beta_j}} \sim t_{n-k-1} \qquad \text{(under } H_0\text{)}$$

2.4.1 The t-ratio Test

- if $\beta_j = c$ is not true the ratio $t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - c}{\hat{\sigma}_{\beta_j}}$ will not be centered around zero and its value will be far from zero indicating a rejection of H_0
- we need a decision rule to determine how far from zero $t_{\hat{\beta}_j}$ should be to reject H_0
- we choose a significance level α for the test - probability of **Type I error**, mistakenly rejecting H_0 when this is true
- customary to choose 10%, 5% or 1% as significance level α

2.4.1 The t-ratio Test

- for $H_0 : \beta_j = c$ against $H_1 : \beta_j \neq c$ we reject the null when $t_{\hat{\beta}_j}$ is well above zero or well below zero - we split the significance level equally between the two tails of the Student-t pdf and reject H_0 if

$$t_{\hat{\beta}_j} < -t^*_{n-k-1,\alpha/2} \qquad \text{or} \qquad t_{\hat{\beta}_j} > t^*_{n-k-1,\alpha/2}$$

  where $t^*_{n-k-1,\alpha/2}$ is the **critical value** associated with a two-sided alternative hypothesis at α% significance level

- for $H_0 : \beta_j = c$ against $H_1 : \beta_j > c$ we reject the null if

$$t_{\hat{\beta}_j} > t^*_{n-k-1,\alpha}$$

- for $H_0 : \beta_j = c$ against $H_1 : \beta_j < c$ we reject the null if

$$t_{\hat{\beta}_j} < -t^*_{n-k-1,\alpha}$$

2.4.1 The t-ratio Test

Rejection Rules for t-ratio Test

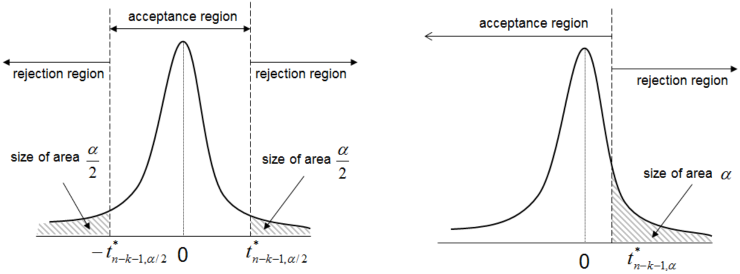
$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - c}{\hat{\sigma}_{\hat{\beta}_j}}$$

two-sided alternative hypothesis

$$H_0 : \beta_j = c$$
$$H_1 : \beta_j \neq c$$

one-sided alternative hypothesis

$$H_0 : \beta_j = c$$
$$H_1 : \beta_j > c$$



see https://shiny.rit.albany.edu/stat/betaprob/ for interactive examples

2.4.1 The t-ratio Test

An Example: House Prices and Interest Rates - continued

Dependent Variable: D_Price
Method: Least Squares
Sample (adjusted): 1973 2007
Included observations: 35 after adjustments
Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.398569	0.541453	2.582990	0.0146
D_Price(-1)	0.747250	0.096000	7.783849	0.0000
D_Rate(-1)	-0.069321	0.045878	-1.510987	0.1406

R-squared	0.476213	Mean dependent var	6.255417
Adjusted R-squared	0.443476	S.D. dependent var	3.440050
S.E. of regression	2.566295	Akaike info criterion	4.804620
Sum squared resid	210.7479	Schwarz criterion	4.937936
Log likelihood	-81.08085	F-statistic	14.54676
Durbin-Watson stat	1.631411	Prob(F-statistic)	0.000032

t-tests for single hypothesis with $H_0 : \beta_j = 0$

2.4.1 The t-ratio Test

- the last column is a p-value or probability associated with the value of the statistic
- p-value is the smallest significance level at which the H_0 can be rejected
- that is, for a two-sided alternative $H_1 : \beta_j \neq 0$

$$p - \text{value} = P(|t_{n-k-1}| > |t_{\hat{\beta}_j}|) = 2P(t_{n-k-1} > t_{\hat{\beta}_j})$$

- higher statistic implies lower p-value and means that there is more evidence in the data to reject the null hypothesis
- as general rule of thumb, if p-value is lower than 5% or 1% we consider there to be enough evidence to safely reject the null

2.4.2 The F-Test

- multiple or joint hypothesis involves several regression coefficients
- for example, $H_0 : \beta_2 = \beta_4 = 0$ involves two restrictions, $\beta_2 = 0$ and $\beta_4 = 0$, and two regression coefficients
- other examples: $H_0 : \beta_2 + \beta_4 = 1$ and $H_0 : \beta_2 = 2, \beta_4 = 0$
- the alternative hypothesis is formulated as the negation of H_0, because several restrictions are involved, it is enough that at least one is false to reject the null
- F-ratio is the statistic to test for a joint hypothesis, to construct the F -ratio, we distinguish between the **unrestricted model** and the **restricted model**
- suppose that we work with the following regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$$

- if we want to test $H_0 : \beta_2 = \beta_4 = 0$, when the null hypothesis is imposed on the unrestricted model above we obtain the restricted model

$$Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + u$$

2.4.2 The F-Test

- after estimating both restricted and unrestricted models we calculate the sum of squared residuals of the unrestricted model SSR_u and the sum of squared residuals of the restricted model SSR_r
- if H_0 is true, estimation results of the unrestricted and restricted models should be very similar, thus SSR_u and SSR_r will not be very different from each other
- if the difference $SSR_r - SSR_u$ is significantly large, we will conclude that there is evidence against H_0

2.4.2 The F-Test

- F-statistic measures statistically the difference in the sum of squared residuals and is defined as

$$F_{m,n-k-1} = \frac{(SSR_r - SSR_u)/m}{SSR_u/(n-k-1)}$$

  where m is the number of restrictions under H_0 and $n-k-1$ is the number of degrees of freedom in the unrestricted model

- $F_{m,n-k-1}$ statistic is distributed as an F random variable with $(m, n-k-1)$ degrees of freedom

- as with the t-ratio, we need to specify a decision rule to reject or fail to reject the null hypothesis - we again choose a significance level α for the test as probability of Type I error, mistakenly rejecting H_0 when this is true

- H_0 is rejected when

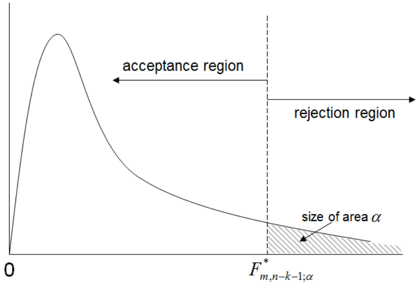
$$F_{m,n-k-1} > F^*_{m,n-k-1;\alpha}$$

  where $F^*_{m,n-k-1;\alpha}$ is the critical value associated with the α% significance level; otherwise we fail to reject the null hypothesis

2.4.2 The F-Test

Rejection Rules for F-Test

$$F_{m,n-k-1} = \frac{(SSR_r - SSR_u)/m}{SSR_u/(n-k-1)}$$



2.4.2 The F-Test

An Example: House Prices and Interest Rates continued

- after an equation has been estimated in EViews, it automatically show the results of the test with $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

Dependent Variable: D_Price
Method: Least Squares
Sample (adjusted): 1973 2007
Included observations: 35 after adjustments
Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.398569	0.541453	2.582990	0.0146
D_Price(-1)	0.747250	0.096000	7.783849	0.0000
D_Rate(-1)	-0.069321	0.045878	-1.510987	0.1406

R-squared	0.476213	Mean dependent var	6.255417
Adjusted R-squared	0.443476	S.D. dependent var	3.440050
S.E. of regression	2.566295	Akaike info criterion	4.804620
Sum squared resid	210.7479	Schwarz criterion	4.937936
Log likelihood	-81.08085	F-statistic	14.54676
Durbin-Watson stat	1.631411	Prob(F-statistic)	0.000032

F-test for overall significance of the regression with joint hypothesis $H_0 : \beta_1 = \beta_2 = 0$

2.4.2 The F-Test

An Example: House Prices and Interest Rates continued

- to perform an F-test with $H_0 : \beta_{j_1} = \beta_{j_2} = \ldots = \beta_{j_l} = 0$ for a subset $\{j_1, j_2, \ldots, j_l\} \subset \{1, 2, \ldots, k\}$ choose **View** $\rightarrow$ **Coefficient Diagnostics** $\rightarrow$ **Redundant Variables Test - Likelihood Ratio** and enter the names of the variables $j_1, j_2, \ldots, j_l$

Redundant Variables Test
Null hypothesis: D_PRICE(-1) D_RATE(-1) are jointly insignificant
Equation: EQ_TBL2_3
Specification: D_PRICE C D_PRICE(-1) D_RATE(-1)
Redundant Variables: D_PRICE(-1) D_RATE(-1)

	Value	df	Probability
F-statistic	14.54676	(2, 32)	0.0000
Likelihood ratio	22.63344	2	0.0000

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	191.6061	2	95.80307
Restricted SSR	402.3540	34	11.83394
Unrestricted SSR	210.7479	32	6.585871

LR test summary:

	Value	df
Restricted LogL	-92.39757	34
Unrestricted LogL	-81.08085	32

2.4.2 The F-Test

An Example: House Prices and Interest Rates continued

- to perform a t-test or an F-test with some other null hypothesis
  - first choose **View** $\rightarrow$ **Representations** to find out which coefficient is which one
  - then choose **View** $\rightarrow$ **Coefficient Diagnostics** $\rightarrow$ **Wald Test - Coefficient Restrictions** and enter the restrictions

Wald Test:
Equation: EQ_TBL2_3

Test Statistic	Value	df	Probability
F-statistic	31.19105	(2, 32)	0.0000
Chi-square	62.38211	2	0.0000

Null Hypothesis: C(2)=0, C(3)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	0.747250	0.096000
C(3)	-0.069321	0.045878

Restrictions are linear in coefficients.