Econometrics

	Course Calendar
Week	Main Content
Week 7	Extension of Simple Regression: Functional Forms I
Week 8	Extension of Simple Regression: Functional Forms II
Week 9	Extension of Simple Regression: Functional Forms III
Week 10	Multiple Regression
Week 11	Multiple Regression: Problem of Inference
Week 12	Multiple Regression: Functional Forms
Week 13	Introduction to Dummy Variables
Week 14	Introduction to Dummy Variables and Regression Methods
Week 15	Regression with Dummy Variables: Hands-on-Exercise
Week 16	Application of Regression

Econometrics

Lecture 15. Dummy Variables in Regression: Hands-on-Exercise

Geetha Rani Prakasam, Ph.D

Professor,

Recap

- Type of Data
- Type of Variables
- Dummy Variables types
- Dummy variables in Multiple regression: ANOVA in example

Outline

- Several applications of dummy variables in Multiple regression in explaining interesting economic aspects with examples
- Regression with a Mixture of Quantitative and Qualitative Regressors: The ANCOVA Models
- **The Dummy Variable Alternative to the Chow Test: Structural Differences in the US Savings and Income data**
- Interaction Effects Using Dummy Variables
- **The Use of Dummy Variables in Seasonal Analysis**

9.4 Regression with a Mixture of Quantitative and Qualitative Regressors: The ANCOVA Models

- Analysis of covariance (ANCOVA) models.
- They are an extension of the ANOVA models in that they provide a method of statistically controlling the effects of quantitative regressors, called **covariates or control variables, in a model that includes** both quantitative and qualitative, or dummy, regressors.
- we develop the following model:
- $Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{2i} + \beta_4 X_i + u_i$ ***(9.4.1)***
- where *Yi = average annual salary of public school teachers in state ($); X_i = spending on public school per pupil ($); D_{2i} = 1, if the state is in the Northeast or North Central* = 0, otherwise; *D_{3i} = 1, if the state is in the South;* = 0, otherwise

9.4 : The ANCOVA Models

- The data on X are given in Table 9.1. Remember, we are treating the West as the benchmark category.
- Also, note that besides the two qualitative regressors, we have a quantitative variable, X, which in the context of the ANCOVA models is known as a **covariate,** as noted earlier.
- From the data in Table 9.1, the results of the model (9.4.1):
- $\hat{Y}_i = 28{,}694.918 - 2{,}954.127D_{2i} - 3{,}112.194D_{3i} + 2.3404X_i$

  se = (3262.521) (1862.576) (1819.873) (0.3592)

  t = (8.795)* (−1.586)** (−1.710)** (6.515)* **(9.4.2)**

  $R^2 = 0.4977$
- where * indicates p values less than 5 %, and ** indicates p values greater than 5 %

9.4 : The ANCOVA Models

- As these results suggest, ceteris paribus: as public expenditure goes up by a dollar, on average, a public school teacher's salary goes up by about $2.34.
- Controlling for spending on education, we now see that the differential intercept coefficient is not significant for either the Northeast and North Central region or for the South.
- These results are different from those of Eq. (9.2.5). But this should not be surprising, for in Eq. (9.2.5) we did not account for the covariate, differences in per pupil public spending on education.

Fig 9.2: Public School Teachers' Salaries by Geographical Region

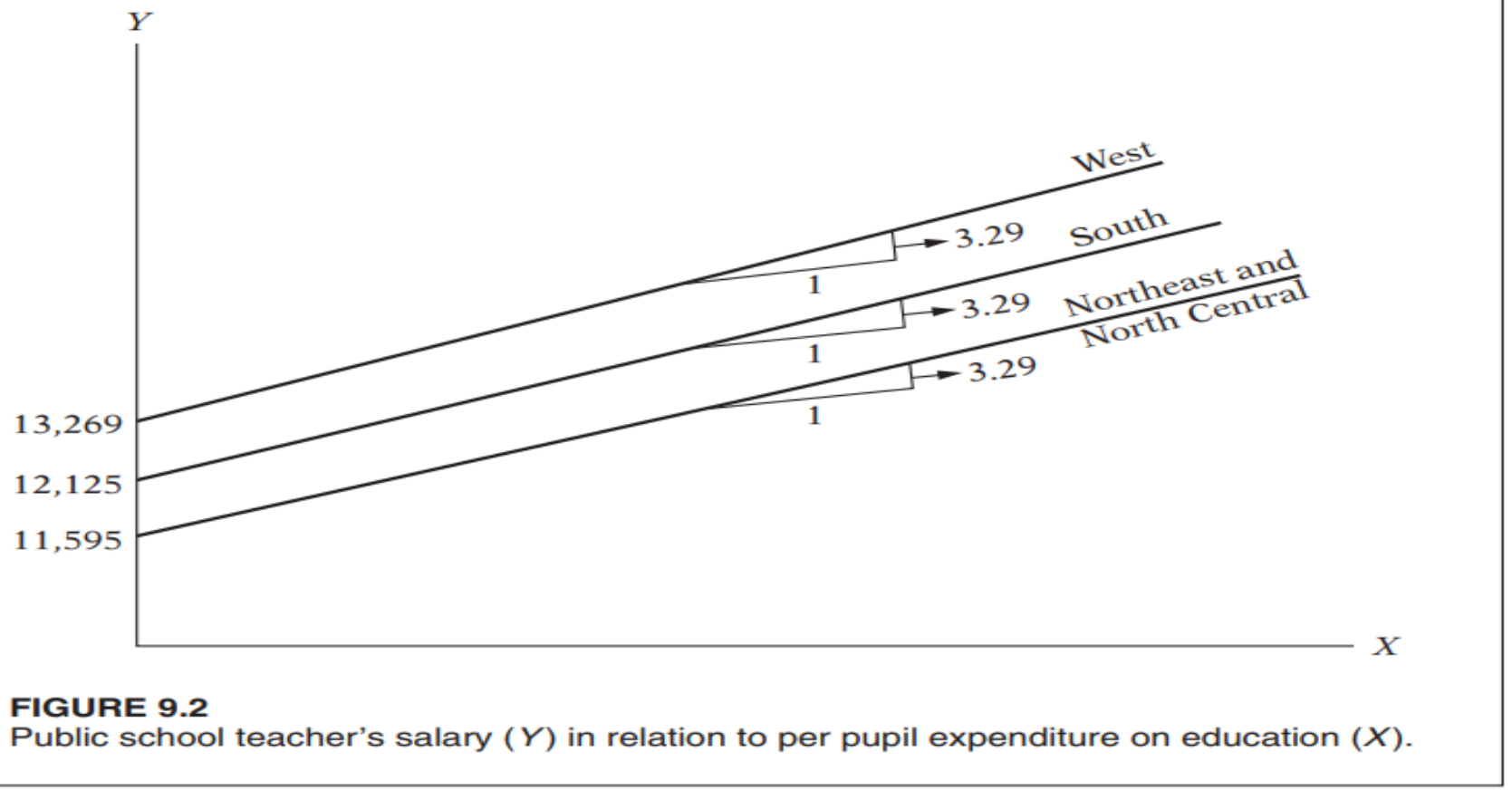


FIGURE 9.2
Public school teacher's salary (Y) in relation to per pupil expenditure on education (X).

Source: Basic Econometrics by Damodar Gujarti, Page. 306

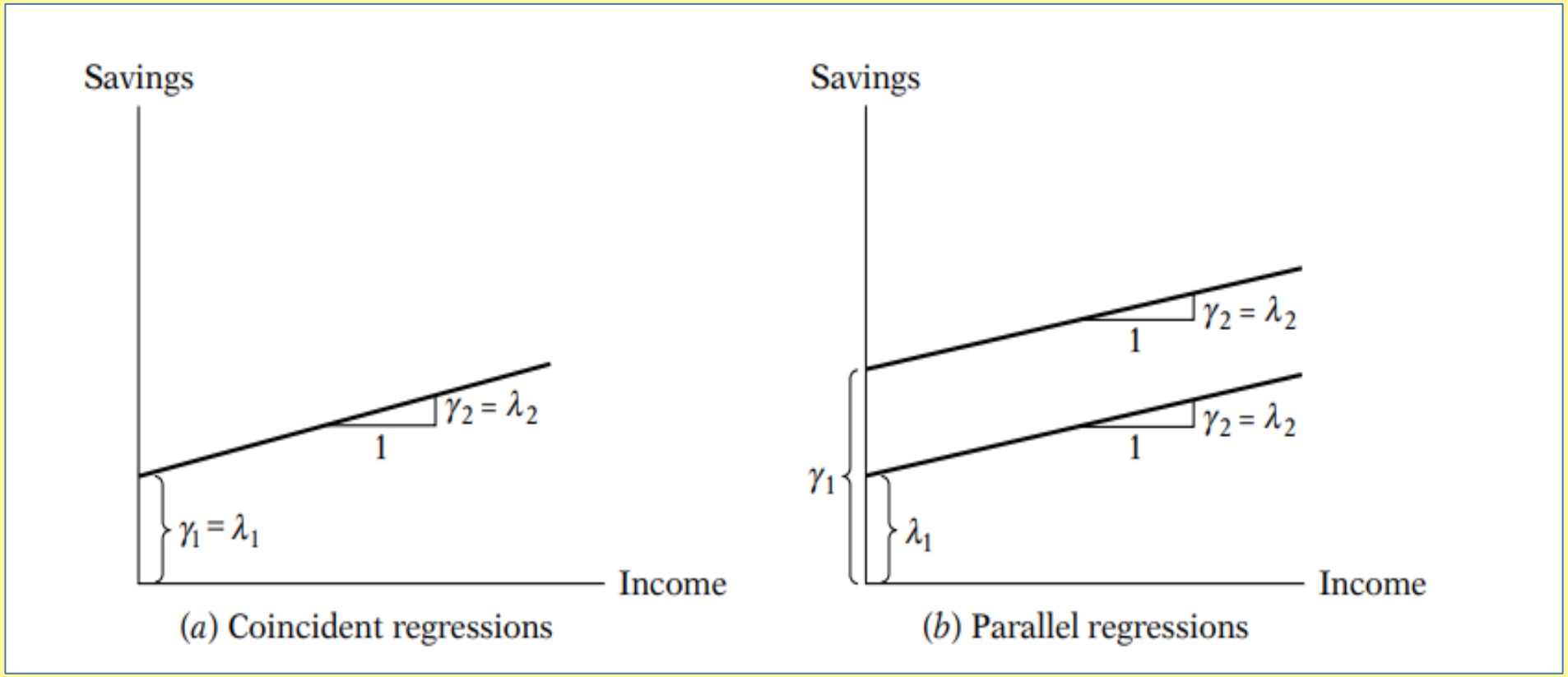
9.4 : The ANCOVA Models

- Diagrammatically, we have the situation shown in Figure 9.2.
- Note that although we have shown three regression lines for the three regions, statistically the regression lines are the same for all three regions.
- Also note that the three regression lines are drawn parallel. (Why?)

9.5 The Dummy Variable Alternative to the Chow Test

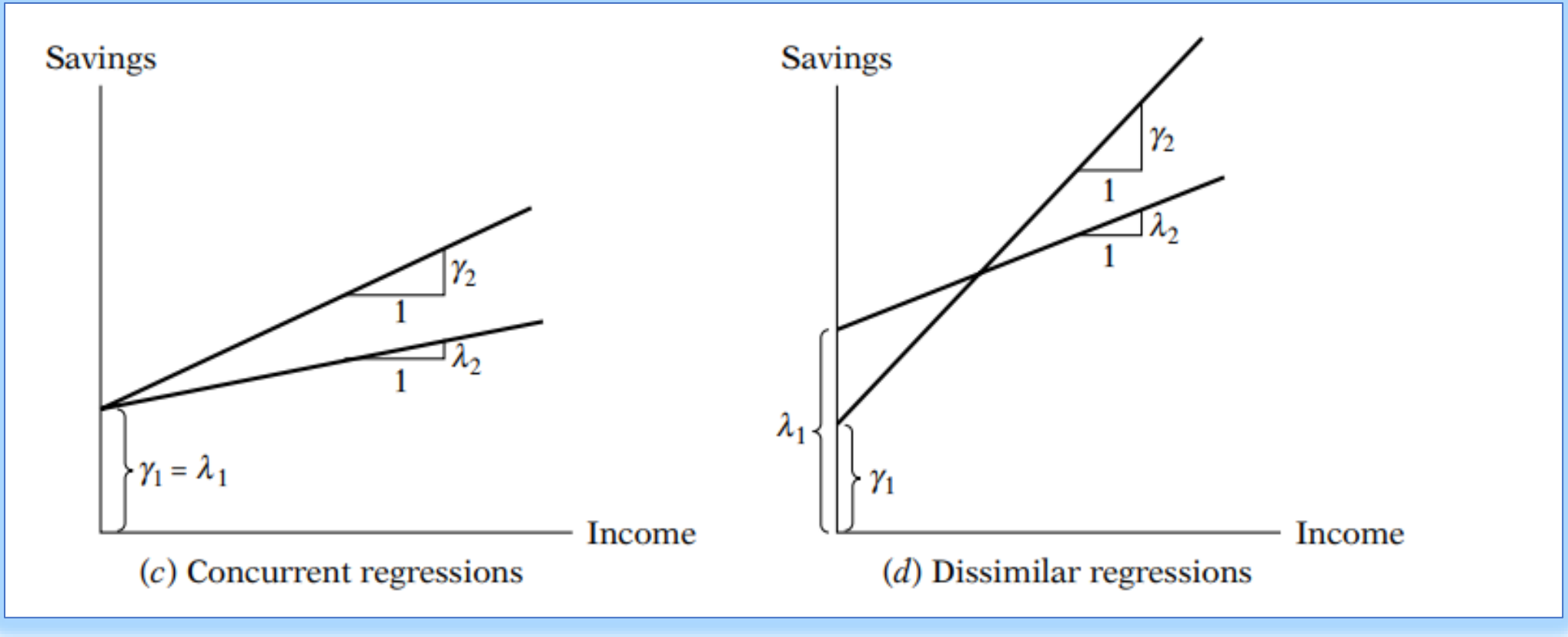
- In lecture 12 we discussed the Chow test to examine the structural stability of a regression model.
- The example on savings and income in the US over the period 1970–1995.
- We divided the sample period into two, 1970–1981 and 1982–1995, and showed on the basis of the Chow test that there was a difference in the regression of savings on income between the two periods.
- we could not tell whether the difference in the two regressions was because of differences in the intercept terms or the slope coefficients or both.
- There are four possibilities, which we illustrate in Figure 9.3 in the next two slides

Fig 9.3: Coincident regressions and Parallel regressions



Source: Basic Econometrics by Damodar Gujarti, Page. 307

Fig 9.3: Concurrent and Dissimilar Regressions



Source: Basic Econometrics by Damodar Gujarti, Page. 307

9.5 The Dummy Variable Alternative to the Chow Test

- 1. Both the intercept and the slope coefficients are the same in the two regressions. This, the case of **coincident regressions, is shown in Figure 9.3*a*.**
- 2. Only the intercepts in the two regressions are different but the slopes are the same. This is the case of **parallel regressions, which is shown in Figure 9.3*b*.**
- 3. The intercepts in the two regressions are the same, but the slopes are different. This is the situation of **concurrent regressions (Figure 9.3*c*).**
- 4. Both the intercepts and slopes in the two regressions are different. This is the case of **dissimilar regressions, which is shown in Figure 9.3*d*.**

9.5 The Dummy Variable Alternative to the Chow Test

- The source of difference, if any, can be pinned down by pooling all the observations (26 in all) and running just one multiple regression as:
- $Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t$ **(9.5.1)**
- $\beta_1 + \beta_2 X_t$
- where Y = savings; X = income; t = time; D = 1 for observations in 1982–1995 = 0, otherwise (i.e., for observations in 1970–1981)
- Table 9.2 shows the structure of the data matrix.
- To see the implications of Eq. (9.5.1), and, assuming, as usual, that E(ui) = 0, we obtain:

Observation	Savings	Income	Dum
1970	61	727.1	0
1971	68.6	790.2	0
1972	63.6	855.3	0
1973	89.6	965	0
1974	97.6	1054.2	0
1975	104.4	1159.2	0
1976	96.4	1273	0
1977	92.5	1401.4	0
1978	112.6	1580.1	0
1979	130.1	1769.5	0
1980	161.8	1973.3	0
1981	199.1	2200.2	0
1982	205.5	2347.3	1
1983	167	2522.4	1
1984	235.7	2810	1
1985	206.2	3002	1
1986	196.5	3187.6	1
1987	168.4	3363.1	1
1988	189.1	3640.8	1
1989	187.8	3894.5	1
1990	208.7	4166.8	1
1991	246.4	4343.7	1
1992	272.6	4613.7	1
1993	214.4	4790.2	1
1994	189.4	5021.7	1
1995	249.3	5320.8	1

Source: Basic Econometrics by Damodar Gujarti, Page. 308

Table 8.9: Savings &Personal Disposable Income (billions $), US, 1970–1995

Observation	Savings	Income	Observation	Savings	Income
1970	61	727.1	1983	167	2522.4
1971	68.6	790.2	1984	235.7	2810
1972	63.6	855.3	1985	206.2	3002
1973	89.6	965	1986	196.5	3187.6
1974	97.6	1054.2	1987	168.4	3363.1
1975	104.4	1159.2	1988	189.1	3640.8
1976	96.4	1273	1989	187.8	3894.5
1977	92.5	1401.4	1990	208.7	4166.8
1978	112.6	1580.1	1991	246.4	4343.7
1979	130.1	1769.5	1992	272.6	4613.7
1980	161.8	1973.3	1993	214.4	4790.2
1981	199.1	2200.2	1994	189.4	5021.7
1982	205.5	2347.3	1995	249.3	5320.8

Source: Economic Report of the President, 1997, Table B-28, p. 332

Basic Econometrics, Damodar Gujarati, Page. 274

9.5 The Dummy Variable Alternative to the Chow Test

- Mean savings function for 1970–1981:
- $E(Y_t \mid D_t = 0, X_t) = \alpha_1 + \beta_1 X_t$ **(9.5.2)**
- Mean savings function for 1982–1995:
- $E(Y_t \mid D_t = 1, X_t) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_t$ **(9.5.3)**
- Notice that these are the same functions as Eqs. (8.7.1) and (8.7.2), with $\lambda_1 = \alpha_1$, $\lambda_2 = \beta_1$, $\gamma_1 = (\alpha_1 + \alpha_2)$, and $\gamma_2 = (\beta_1 + \beta_2)$.
- Therefore, estimating Eq. (9.5.1) is equivalent to estimating the two individual savings functions in Eqs. (8.7.1) and (8.7.2).

9.5 The Dummy Variable Alternative to the Chow Test

- In Eq. (9.5.1), α_2 is the **differential intercept,** as previously, and **β_2 is the differential slope coefficient (also called the slope drifter),** indicating by how much the slope coefficient of the second period's savings function (the category that receives the dummy value of 1) differs from that of the first period.
- Notice how the introduction of the dummy variable D in the **interactive, or multiplicative, form (D multiplied by X) enables us to differentiate** between slope coefficients of the two periods, just as the introduction of the dummy variable in the **additive form enabled us to distinguish between the intercepts of** the two periods.

EX 9.4: Structural Differences in the U.S. Savings– Income Regression, DV Approach

- First present the regression results of model (9.5.1) applied to the U.S. savings–income data.
- $\hat{Y}_t = 1.0161 + 152.4786D_t + 0.0803X_t - 0.0655(D_tX_t)$

  se = (20.1648) (33.0824) (0.0144) (0.0159) **(9.5.4)**

  t = (0.0504)** (4.6090)* (5.5413)* (–4.0963)*

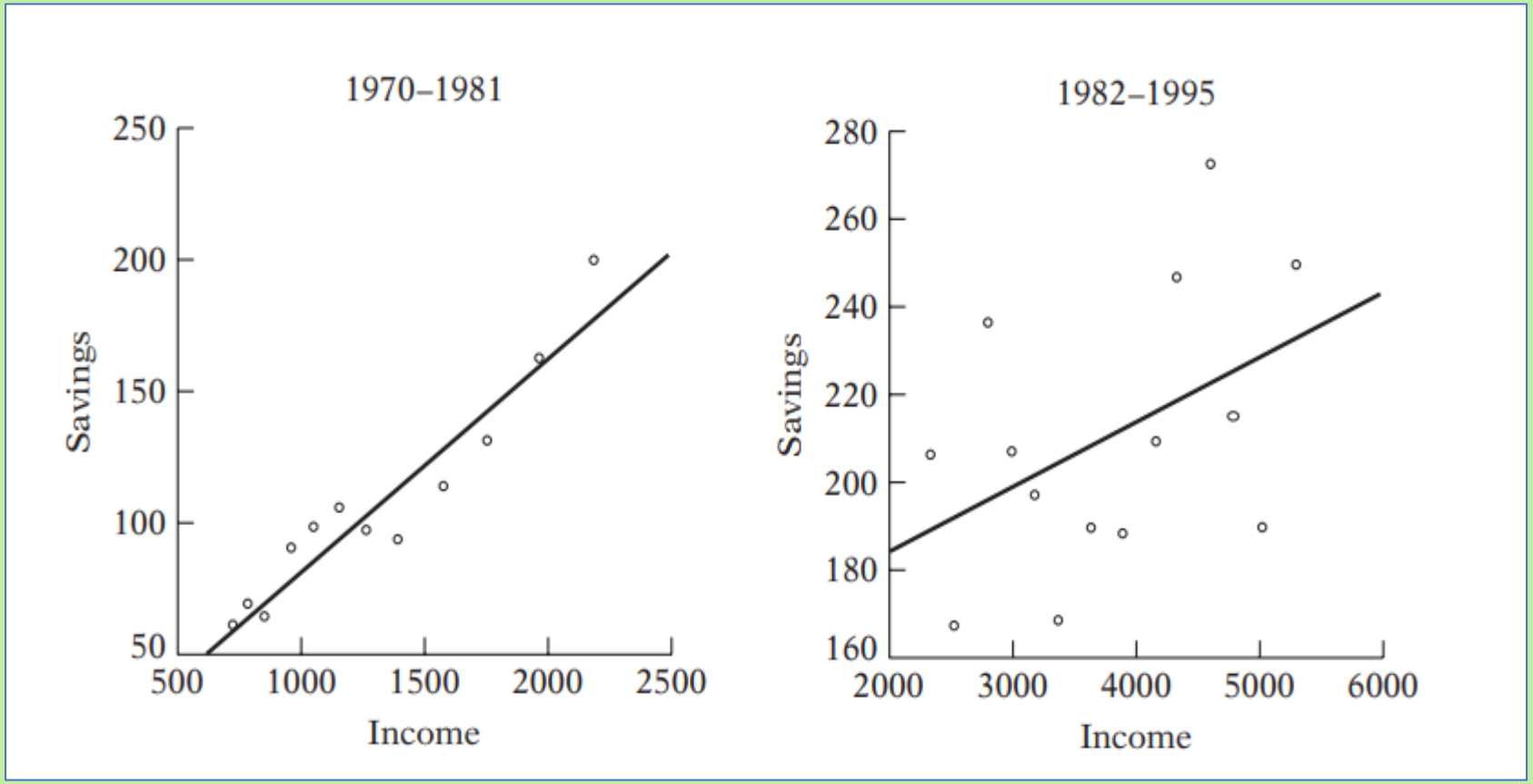
  $R^2 = 0.8819$
- where * indicates p values less than 5 % and ** indicates p values greater than 5 %.
- As these regression results show, both the differential intercept and slope coefficients are statistically significant, strongly suggesting that the savings–income regressions for the two time periods are different, as in Figure 9.3*d.*

EX 9.4 Structural Differences in the U.S. Savings– Income Regression, DV Approach

- From Eq. (9.5.4), we can derive equations (9.5.2) and (9.5.3), which are: Savings–income regression, 1970–1981:
- $\hat{Y}_t = 1.0161 + 0.0803X_t$ **(9.5.5)**
- Savings–income regression, 1982–1995:
- $\hat{Y}_t = (1.0161 + 152.4786) + (0.0803 - 0.0655)X_t$

  $= 153.4947 + 0.0148X_t$ **(9.5.6)**
- These are precisely the results we obtained in Eqs. (8.7.1a) and (8.7.2a), which should not be surprising.
- These regressions are already shown in Figure 8.3.

Figure 8.3: U.S. Savings– Income Regression



Source: Basic Econometrics by Damodar Gujarti, Page. 276

EX 9.4 Structural Differences in the U.S. Savings– Income Regression, DV Approach

- The advantages of the dummy variable technique (i.e., estimating Eq. [9.5.1]) over the Chow test (i.e., estimating the three regressions [8.7.1], [8.7.2], and [8.7.3]) can now be seen readily:
- We need to run only a single regression because the individual regressions can easily be derived from it in the manner indicated by equations (9.5.2) and (9.5.3).
- The single regression (9.5.1) can be used to test a variety of hypotheses.
- Thus if the differential intercept coefficient α_2 is statistically insignificant, we may accept the hypothesis that the two regressions have the same intercept, that is, the two regressions are concurrent (see Figure 9.3c).

EX 9.4 Structural Differences in the U.S. Savings– Income Regression, DV Approach

- Similarly, if the differential slope coefficient β_2 is statistically insignificant but α_2 is significant, we may not reject the hypothesis that the two regressions have the same slope, that is, the two regression lines are parallel (cf. Figure 9.3b).
- The test of the stability of the entire regression (i.e., $\alpha = \beta_2 = 0$, simultaneously) can be made by the usual F test (recall the restricted least-squares F test).
- If this hypothesis is not rejected, the regression lines will be coincident, as shown in Figure 9.3a.

EX 9.4 Structural Differences in the U.S. Savings– Income Regression, DV Approach

- The Chow test does not explicitly tell us *which coefficient, intercept, or slope is* different, or whether (as in this example) both are different in the two periods.
- *We cannot tell, via the* Chow test, which one of the four possibilities depicted in Figure 9.3 exists in a given instance.
- Finally, since pooling (i.e., including all the observations in one regression) increases the degrees of freedom, it may improve the relative precision of the estimated parameters.
- Of course, keep in mind that every addition of a dummy variable will consume one degree of freedom.

9.6 Interaction Effects Using Dummy Variables

- DVs are a flexible tool - can handle a variety of interesting problems.
- $Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{31} + \beta X_i + u_i$ **(9.6.1)**
- where Y = hourly wage in dollars;
- X = education (years of schooling);
- D_2 = 1 if female, 0 otherwise;
- D_3 = 1 if nonwhite and non-Hispanic, 0 otherwise
- Gender & race are qualitative regressors; education is a quantitative var.
- Implicit assumption is that the differential effect of the gender dummy D_2 is constant across the two categories of race and the differential effect of the race dummy D_3 is also constant across the two gender types.

9.6 Interaction Effects Using Dummy Variables

- If the mean salary is higher for males than for females, this is so whether they are nonwhite/non-Hispanic or not.
- Likewise, if, say, nonwhite/non-Hispanics have lower mean wages, this is so whether they are females or males.
- In many applications such an assumption may be untenable. A female nonwhite/ non-Hispanic may earn lower wages than a male nonwhite/non-Hispanic.
- In other words, there may be **interaction between the two qualitative variables D_2 and D_3.**

9.6 Interaction Effects Using Dummy Variables

- Therefore their effect on mean Y may not be simply additive as in Eq. (9.6.1) but multiplicative as well, as in the following model.
- $Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + u_i$ **(9.6.2)**
  where the variables are as defined for model (9.6.1).
- From Eq. (9.6.2), we obtain:
- $E(Y_i \mid D_{2i} = 1, D_{3i} = 1, X_i) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i$ **(9.6.3)**
- which is the mean hourly wage function for female nonwhite/non-Hispanic workers.
- Note: α_2 = differential effect of being a female

9.6 Interaction Effects Using Dummy Variables

- α_3 = differential effect of being a nonwhite/non-Hispanic
- α_4 = differential effect of being a female nonwhite/non-Hispanic
- which shows that the mean hourly wages of female nonwhite/non-Hispanics is different (by α_4) from the mean hourly wages of females or nonwhite/non-Hispanics.
- If, for instance, all three differential dummy coefficients are negative, this would imply that female nonwhite/non-Hispanic workers earn much lower mean hourly wages than female or nonwhite/non-Hispanic workers as compared with the base category, which in the present example is male white or Hispanic.

9.6 Interaction Effects Using Dummy Variables

- We can see how the **interaction dummy** (i.e., the product of two qualitative or DV variables) modifies the effect of the two attributes considered individually (i.e., additively).
- Regression results based on model (9.6.1), using the data in table, we get:
- $\hat{Y}_i = -0.2610 - 2.3606D_{2i} - 1.7327D_{3i} + 0.8028X_i$
  $t = (-0.2357)^{**}\ (-5.4873)^{*}\ (-2.1803)^{*}\ (9.9094)^{*}$ **(9.6.4)**
  $R^2 = 0.2032;\ n = 528$
- where * indicates p values less than 5 % and ** indicates p values greater than 5 %.

EX 9.5 Average Hourly Earnings in Relation to Education, Gender, & Race

- We can check that the differential intercept coefficients are statistically significant, that they have the expected signs (why?), and that education has a strong positive effect on hourly wage, an unsurprising finding.
- As Eq. (9.6.4) shows, ceteris paribus, the average hourly earnings of females are lower by about $2.36, and the average hourly earnings of nonwhite non-Hispanic workers are also lower by about $1.73.

EX 9.5 Average Hourly Earnings in Relation to Education, Gender, and Race

- Consider the results of model (9.6.2), which includes the interaction dummy.
- $\hat{Y}_i = -0.26100 - 2.3606D_{2i} - 1.7327D_{3i} + 2.1289D_{2i}D_{3i} + 0.8028X_i$
  t = (−0.2357)** (−5.4873)* (−2.1803)* (1.7420)** (9.9095)** **(9.6.5)**
  $R^2 = 0.2032$; n = 528
- where * indicates p values less than 5 % &** indicates p values greater than 5 %.
- As we can see, the two additive dummies are still statistically significant, but the interactive dummy is not at the conventional 5 percent level; the actual p value of the interaction dummy is about the 8 percent level.
- If this is a low enough probability, then the results of Eq. (9.6.5) can be interpreted as follows:

EX 9.5 Average Hourly Earnings in Relation to Education, Gender, and Race

- Holding the level of education constant, if we add the three dummy coefficients we will obtain: −1.964 (=−2.3605 − 1.7327 + 2.1289), which means that mean hourly wages of nonwhite/non-Hispanic female workers is lower by about $1.96, which is between the value of −2.3605 (gender difference alone) and −1.7327 (race difference alone).
- The preceding example clearly reveals the role of interaction dummies when two or more qualitative regressors are included in the model.
- It is important to note that in the model (9.6.5) we are assuming that the rate of increase of hourly earnings with respect to education (of about 80 cents per additional year of schooling) remains constant across gender and race.
- But this may not be the case. If we want to test for this, we will have to introduce differential slope coefficients (see Ex. 9.25).

9.7: The Use of Dummy Variables in Seasonal Analysis

- Many economic time series based on monthly or quarterly data exhibit seasonal patterns.
- Often it is desirable to remove the seasonal factor, or *component, from a time series so that one can concentrate on the other components,* such as the trend.
- A time series may contain four components: (1) **seasonal, (2) cyclical, (3) trend, and (4) strictly** random.
- The process of removing the seasonal component from a time series is known as **deseasonalization or seasonal adjustment, and the time series thus** obtained is called the **deseasonalized, or seasonally adjusted, time series.**

9.7: The Use of Dummy Variables in Seasonal Analysis

- **I**mportant economic time series, such as the unemployment rate, the consumer price index (CPI), the producer's price index (PPI), and the index of industrial production, are usually published in seasonally adjusted form. (regular oscillatory movements).
- There are several methods of deseasonalizing a time series, but we will consider only one of these methods, namely, the method of dummy variables.
- Table 9.3 gives quarterly data for the years 1978–1995 on the sale of four major appliances, dishwashers, garbage disposers, refrigerators, and washing machines, all data in thousands of units.

Table 9.3: Quarterly Data on Appliance Sales (in thousands) and Expenditure on Durable Goods

DISH	DISP	FRIG	WASH	DUR	DISH	DISP	FRIG	WASH	DUR
841	798	1317	1271	252.6	480	706	943	1036	247.7
957	837	1615	1295	272.4	530	582	1175	1019	249.1
999	821	1662	1313	270.9	557	659	1269	1047	251.8
960	858	1295	1150	273.9	602	837	973	918	262
894	837	1271	1289	268.9	658	867	1102	1137	263.3
851	838	1555	1245	262.9	749	860	1344	1167	280
863	832	1639	1270	270.9	827	918	1641	1230	288.5
878	818	1238	1103	263.4	858	1017	1225	1081	300.5
792	868	1277	1273	260.6	808	1063	1429	1326	312.6
589	623	1258	1031	231.9	840	955	1699	1228	322.5
657	662	1417	1143	242.7	893	973	1749	1297	324.3
699	822	1185	1101	248.6	950	1096	1117	1198	333.1
675	871	1196	1181	258.7	838	1086	1242	1292	344.8
652	791	1410	1116	248.4	884	990	1684	1342	350.3
628	759	1417	1190	255.5	905	1028	1764	1323	369.1
529	734	919	1125	240.4	909	1003	1328	1274	356.4

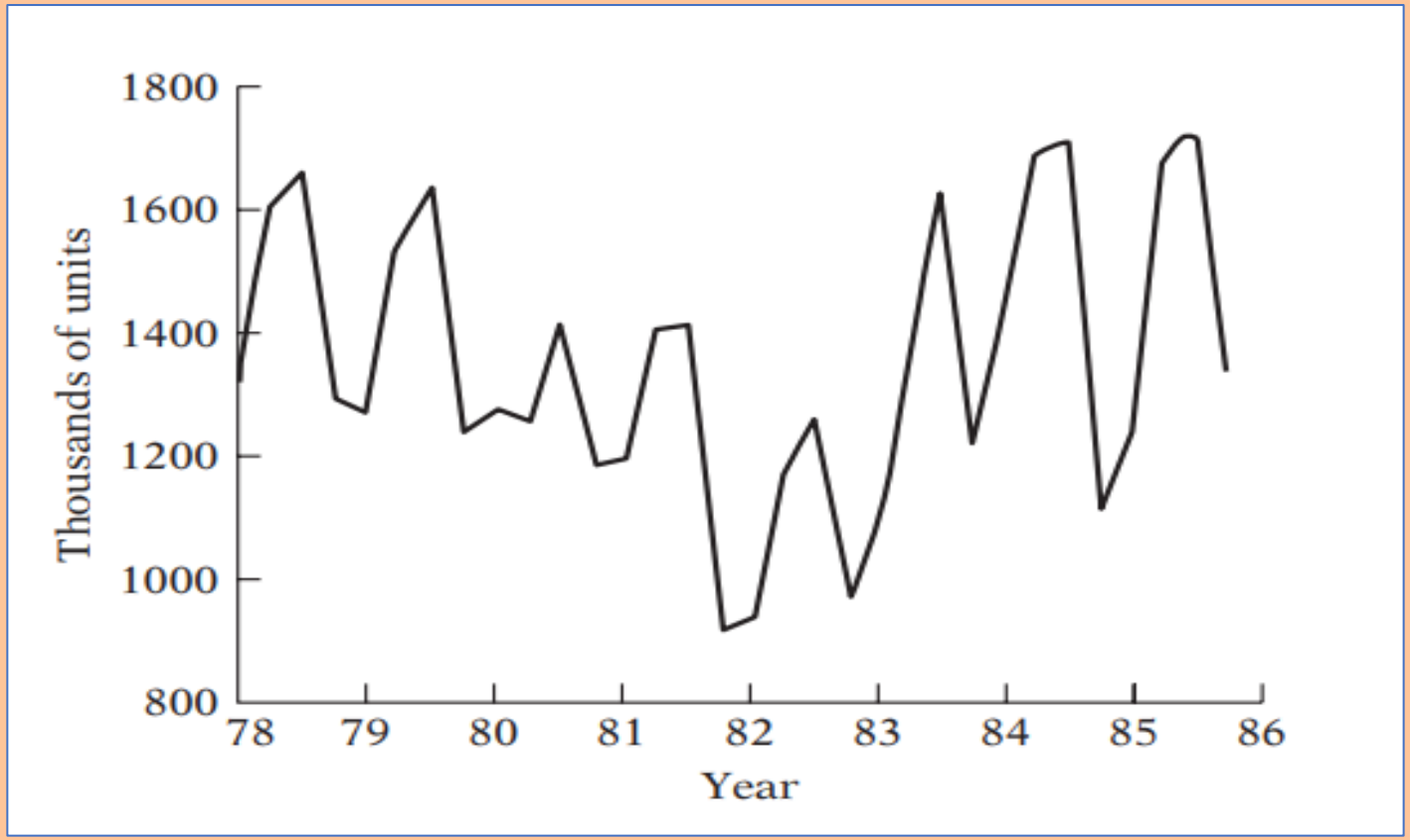
Note: DISH = dishwashers; DISP = garbage disposers; FRIG = refrigerators; WASH = dishwashers; DUR = durable goods expenditure, billions of 1992 dollars. Source: Business Statistics and Survey of Current Business, Department of Commerce (various issues)

Source: Basic Econometrics, by Damodar Gujarti, Page. 313

9.7 The Use of Dummy Variables in Seasonal Analysis

- It also gives data on durable goods expenditure in 1982 billions of dollars.
- To illustrate the dummy technique, we will consider only the sales of refrigerators over the sample period.
- This data shown in Figure 9.4 suggests that perhaps there is a seasonal pattern in the data associated with various quarters.

FIGURE 9.4 Sales of refrigerators 1978–1985 (quarterly)



Source: Basic Econometrics, by Damodar Gujarti, Page. 313

TABLE 9.4 U.S. Refrigerator Sales (in 000s), 1978–1995 (Quarterly)

FRIG	DUR	D_2	D_3	D_4	FRIG	DUR	D_2	D_3	D_4
1317	252.6	0	0	0	943	247.7	0	0	0
1615	272.4	1	0	0	1175	249.1	1	0	0
1662	270.9	0	1	0	1269	251.8	0	1	0
1295	273.9	0	0	1	973	262.0	0	0	1
1271	268.9	0	0	0	1102	263.3	0	0	0
1555	262.9	1	0	0	1344	280.0	1	0	0
1639	270.9	0	1	0	1641	288.5	0	1	0
1238	263.4	0	0	1	1225	300.5	0	0	1
1277	260.6	0	0	0	1429	312.6	0	0	0
1258	231.9	1	0	0	1699	322.5	1	0	0
1417	242.7	0	1	0	1749	324.3	0	1	0
1185	248.6	0	0	1	1117	333.1	0	0	1
1196	258.7	0	0	0	1242	344.8	0	0	0
1410	248.4	1	0	0	1684	350.3	1	0	0
1417	255.5	0	1	0	1764	369.1	0	1	0
919	240.4	0	0	1	1328	356.4	0	0	1

Source: Basic Econometrics, by Damodar Gujarti, Page. 314

9.7 The Use of Dummy Variables in Seasonal Analysis

- To see if this is the case, consider the following model:
- $Y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + u_t$ **(9.7.1)**
- where Y_t = sales of refrigerators (in thousands) and the D's are the dummies, taking a value of 1 in the relevant quarter and 0 otherwise.
- To avoid the DV trap, we are assigning a dummy to each quarter of the year, but omitting the intercept term.
- If there is any seasonal effect in a given quarter, that will be indicated by a statistically significant t value of the dummy coefficient for that quarter.

9.7 The Use of Dummy Variables in Seasonal Analysis

- Notice that in Eq. (9.7.1) we are regressing Y effectively on an intercept, except that we allow for a different intercept in each season (i.e., quarter).
- As a result, the dummy coefficient of each quarter will give us the mean refrigerator sales in each quarter or season.
- From the data on refrigerator sales in Table 9.4, we get:
- $\hat{Y}_t = 1{,}222.125D_{1t} + 1{,}467.500D_{2t} + 1{,}569.750D_{3t} + 1{,}160.000D_{4t}$

  t = (20.3720) (24.4622) (26.1666) (19.3364) **(9.7.2)**

  $R^2 = 0.5317$

EX. 9.6: Seasonality in Refrigerator Sales

- Note: We have not given the standard errors of the estimated coefficients, as each standard error is equal to 59.9904, because all the dummies take only a value of 1 or zero.
- The estimated α coefficients in Eq. (9.7.2) represent the average, or mean, sales of refrigerators (in thousands of units) in each season (i.e., quarter).
- Thus, the average sale of refrigerators in the first quarter, in thousands of units, is about 1,222, that in the second quarter about 1,468, that in the third quarter about 1,570, and that in the fourth quarter about 1,160.

EX. 9.6: Seasonality in Refrigerator Sales

- Instead of assigning a dummy for each quarter and suppressing the intercept term to avoid the DV trap, we could assign three dummies and include the intercept term. And we get :
- $\hat{Y}_t = 1{,}222.1250 + 245.3750D_{2t} + 347.6250D_{3t} - 62.1250D_{4t}$

  $t = (20.3720)^* \ (2.8922)^* \ (4.0974)^* \ (-0.7322)^{**}$ **(9.7.3)**

  $R^2 = 0.5318$;

  where * indicates p values less than 5 %** indicates p values greater than 5 %.
- As we are treating the first quarter as the benchmark, the coefficients attached to the various dummies are now **differential intercepts**, showing by how much the average value of Y in the quarter that receives a dummy value of 1 differs from that of the benchmark quarter.

EX. 9.6: Seasonality in Refrigerator Sales

- Put differently, the coefficients on the seasonal dummies will give the seasonal increase or decrease in the average value of Y relative to the base season.
- If we add the various differential intercept values to the benchmark average value of 1,222.125, we will get the average value for the various quarters.
- we can reproduce exactly Eq. (9.7.2), except for the rounding errors.
- But now we will see the value of treating one quarter as the benchmark quarter, for Eq. (9.7.3) shows that the average value of Y for the fourth quarter is not statistically different from the average value for the first quarter, as the dummy coefficient for the fourth quarter is not statistically significant.

EX. 9.6: Seasonality in Refrigerator Sales

- Our answer will change, depending on which quarter we treat as the benchmark quarter, but the overall conclusion will not change.
- How do we obtain the deseasonalized time series of refrigerator sales?
- we estimate the values of Y from model (9.7.2) (or [9.7.3]) for each observation and subtract them from the actual values of Y, that is, we obtain ($Y_t - \hat{Y}_t$) which are simply the residuals from the regression (9.7.2). We show them in Table 9.5.(next slide)
- To these residuals, we have to add the mean of the Y series to get the forecasted values.

	Actual	Fitted	Residuals	Residual graph 0
1978-I	1317	1222.12	94.875	
1978-II	1615	1467.50	147.500	
1978-III	1662	1569.75	92.250	
1978-IV	1295	1160.00	135.000	
1979-I	1271	1222.12	48.875	
1979-II	1555	1467.50	87.500	
1979-III	1639	1569.75	69.250	
1979-IV	1238	1160.00	78.000	
1980-I	1277	1222.12	54.875	
1980-II	1258	1467.50	−209.500	
1980-III	1417	1569.75	−152.750	
1980-IV	1185	1160.00	25.000	
1981-I	1196	1222.12	−26.125	
1981-II	1410	1467.50	−57.500	
1981-III	1417	1569.75	−152.750	
1981-IV	919	1160.00	−241.000	
1982-I	943	1222.12	−279.125	
1982-II	1175	1467.50	−292.500	
1982-III	1269	1569.75	−300.750	
1982-IV	973	1160.00	−187.000	
1983-I	1102	1222.12	−120.125	
1983-II	1344	1467.50	−123.500	
1983-III	1641	1569.75	71.250	
1983-IV	1225	1160.00	65.000	
1984-I	1429	1222.12	206.875	
1984-II	1699	1467.50	231.500	
1984-III	1749	1569.75	179.250	
1984-IV	1117	1160.00	−43.000	
1985-I	1242	1222.12	19.875	
1985-II	1684	1467.50	216.500	
1985-III	1764	1569.75	194.250	
1985-IV	1328	1160.00	168.000	
				− 0 +

Table 9.5:REFRIGERATOR SALES REGRESSION: ACTUAL, FITTED, AND RESIDUAL VALUES (EQ. 9.7.3)

Source: Basic Econometrics, Damodar Gujarti, Page, 316

EX. 9.6: Seasonality in Refrigerator Sales

- What do these residuals represent?
- They represent the remaining components of the refrigerator time series, namely, the trend, cycle, and random components .
- But, this assumes that the DV technique is an appropriate method of deseasonalizing a time series and that a time series (TS) can be represented as:
- TS = s + c + t + u, where s represents the seasonal, t the trend, c the cyclical, and u the random component.
- Since models (9.7.2) and (9.7.3) do not contain any covariates, will the picture change if we bring in a quantitative regressor in the model?

EX. 9.6: Seasonality in Refrigerator Sales

- Since expenditure on durable goods has an important factor influence on the demand for refrigerators, let us expand our model (9.7.3) by bringing in this variable.
- The data for durable goods expenditure in billions of 1982 dollars are already given in Table 9.3. This is our (quantitative) X variable in the model.
- $\hat{Y}_t = 456.2440 + 242.4976D_{2t} + 325.2643D_{3t} - 86.0804D_{4t} + 2.7734X_t$

  t = (2.5593)* (3.6951)* (4.9421)* (−1.3073)** (4.4496)* **(9.7.4)**

  $R^2 = 0.7298$;

  where * indicates p values less than 5 %** indicates p values greater than 5 %.

EX. 9.6: Seasonality in Refrigerator Sales

- Treating the first quarter as our base, as in Eq. (9.7.3), the differential intercept coefficients for the second and third quarters are statistically different from that of the first quarter, but the intercepts of the fourth quarter and the first quarter are statistically about the same.
- The coefficient of X (durable goods expr.) of about 2.77 tells that, allowing for seasonal effects, if expr. on durable goods goes up by a dollar, on average, sales of refrigerators go up by about 2.77 units, approximately 3 units;
- To bear in mind that refrigerators are in thousands of units and X is in (1982) billions of dollars.

EX. 9.6: Seasonality in Refrigerator Sales

- An interesting question here is:
- Just as sales of refrigerators exhibit seasonal patterns, would not expr. on durable goods also exhibit seasonal patterns?
- How then do we take into account seasonality in X?
- The interesting thing about Eq. (9.7.4) is that the DVs in that model not only remove the seasonality in Y but also the seasonality, if any, in X.
- This follows from a well-known theorem in statistics, known as the **Frisch–Waugh theorem.**
- So to speak, we kill (deseasonalize) two birds (two series) with one stone (the dummy technique).

Seasonality in Refrigerator Sales: Steps

- If we want an informal proof of the preceding statement, just follow these steps:
- (1) Run the regression of Y on the dummies as in Eq. (9.7.2) or Eq. (9.7.3) and save the residuals, say, S1; these residuals represent deseasonalized Y.
- (2) Run a similar regression for X and obtain the residuals from this regression, say, S2; these residuals represent deseasonalized X.
- (3) Regress S1 on S2. we will find that the slope coefficient in this regression is precisely the coefficient of X in *the regression (9.7.4).*

Summary and Conclusions

- 1. Dummy variables, taking values of 1 and zero (or their linear transforms), are a means of introducing qualitative regressors in regression models.
- 2. Dummy variables are a data-classifying device in that they divide a sample into various subgroups based on qualities or attributes (gender, marital status, race, religion, etc.) and *implicitly allow one to run individual regressions for each subgroup.*
- *If there are* differences in the response of the regressand to the variation in the qualitative variables in the various subgroups, they will be reflected in the differences in the intercepts or slope coefficients, or both, of the various subgroup regressions.

Summary and Conclusions

- 3. DV technique needs to be handled carefully. First, if the regression contains a constant term, the number of DVs must be m-1.
- Second, the coefficient attached to the dummy variables must always be interpreted in relation to the base, or reference, group—i.e, the group that receives the value of zero. The base chosen will depend on the purpose at hand.
- Finally, if a model has several qualitative variables with several classes, introduction of DV can consume a large number of df.
- 4. Among its various applications, this lecture considered: (1) comparing two (or more) regressions, (2) deseasonalizing time series data, & (3) interactive dummies,

Reference

- Chapter 9: Dummy Variable Regression Models, in Basic Econometrics by Damodar Gujarati.

What next?

- various other DV applications (a) interpretation of dummies in semilog models, and (b) piecewise linear regression models.
- Applications with Examples