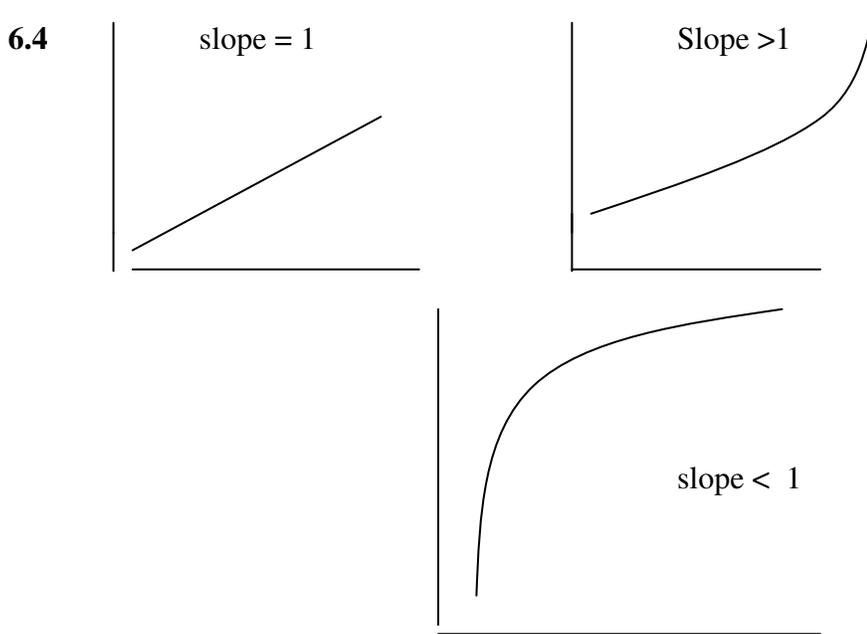


**CHAPTER 6:  
EXTENSIONS OF THE TWO-VARIABLE LINEAR REGRESSION  
MODEL**

- 6.1** *True.* Note that the usual OLS formula to estimate the intercept is  

$$\hat{\beta}_1 = (\text{mean of the regressand} - \hat{\beta}_2 \text{ mean of the regressor}).$$
 But when Y and X are in deviation form, their mean values are always zero. Hence in this case the estimated intercept is also zero.
- 6.2** (a) & (b) In the first equation an intercept term is included. Since the intercept in the first model is not statistically significant, say at the 5% level, it may be dropped from the model.
- (c) For each model, a one percentage point increase in the monthly market rate of return lead on average to about 0.76 percentage point increase in the monthly rate of return on Texaco common stock over the sample period.
- (d) As discussed in the chapter, this model represents the *characteristic line* of investment theory. In the present case the model relates the monthly return on the Texaco stock to the monthly return on the market, as represented by a broad market index.
- (e) No, the two  $r^2$ s are not comparable. The  $r^2$  of the interceptless model is the raw  $r^2$ .
- (f) Since we have a reasonably large sample, we could use the Jarque-Bera test of normality. The JB statistic for the two models is about the same, namely, 1.12 and the *p value* of obtaining such a JB value is about 0.57. Hence do not reject the hypothesis that the error terms follow a normal distribution.
- (g) As per Theil's remark discussed in the chapter, if the intercept term is absent from the model, then running the regression through the origin will give more efficient estimate of the slope coefficient, which it does in the present case.
- 6.3** (a) Since the model is linear in the parameters, it is a linear regression model.
- (b) Define  $Y^* = (1/Y)$  and  $X^* = (1/X)$  and do an OLS regression of  $Y^*$  on  $X^*$ .
- (c) As X tends to infinity, Y tends to  $(1/\beta_1)$ .

(d) Perhaps this model may be appropriate to explain low consumption of a commodity when income is large, such as an inferior good.



6.5 For Model I we know that

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}, \text{ where X and Y are in deviation form.}$$

For Model II, following similar step, we obtain:

$$\hat{\alpha}_2 = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}} = \frac{\sum (x_i / S_x)(y_i / S_y)}{\sum (x_i / S_x)^2} = \frac{\sum (x_i y_i) / S_x S_y}{\sum x_i^2 / S_x^2} =$$

$$\frac{S_x \sum x_i y_i}{S_y \sum x_i^2} = \frac{S_x}{S_y} \hat{\beta}_2$$

This shows that the slope coefficient is not invariant to the change of scale.

6.6 We can write the first model as:

$$\ln(w_1 Y_i) = \alpha_1 + \alpha_2 \ln(w_2 X_i) + u_i^*, \text{ that is,}$$

$\ln w_1 + \ln Y_i = \alpha_1 + \alpha_2 \ln w_2 + \alpha_2 \ln X_i + u_i^*$ , using properties of the logarithms. Since the  $w$ 's are constants, collecting terms, we can simplify this model as:

$$\ln Y_i = (\alpha_1 + \alpha_2 \ln w_2 - \ln w_1) + \alpha_2 X_i + u_i^*$$

$$= A + \alpha_2 \ln X_i + u_i^*$$

where  $A = (\alpha_1 + \alpha_2 \ln w_2 - \ln w_1)$

Comparing this with the second model, you will see that except for the intercept terms, the two models are the same. Hence the estimated slope coefficients in the two models will be the same, the only difference being in the estimated intercepts.

(b) The  $r^2$  values of the two models will be the same.

- 6.7** Equation (6.6.8) is a growth model, whereas (6.6.10) is a linear trend model. The former gives the relative change in the regressand, whereas the latter gives the absolute change. For comparative purposes it is the relative change that may be more meaningful.
- 6.8** The null hypothesis is that the true slope coefficient is 0.005. The alternative hypothesis could be one or two-sided. Suppose we use the two-sided alternative. The estimated slope value is 0.00705. Using the  $t$  test, we obtain:

$$t = \frac{0.00705 - 0.005}{0.00018} = 11.3889$$

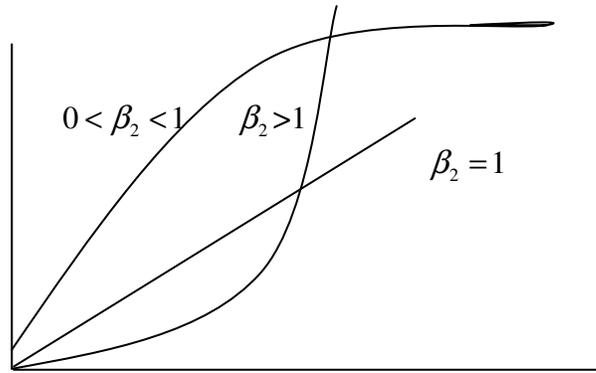
This  $t$  is highly significant. We can therefore reject the null hypothesis.

- 6.9** This can be obtained approximately as:  $18.5508/3.2514 = 5.7055$ , percent.
- 6.10** As discussed in Sec. 6.7 of the text, for most commodities the Engel model depicted in Fig. 6.6(c) seems appropriate. Therefore, the second model given in the exercise may be the choice.
- 6.11** As it stands, the model is not linear in the parameter. But consider the following “trick.” First take the ratio of  $Y$  to  $(1-Y)$  and then take the natural log of the ratio. This transformation will make the model linear in the parameters. That is, run the following regression:

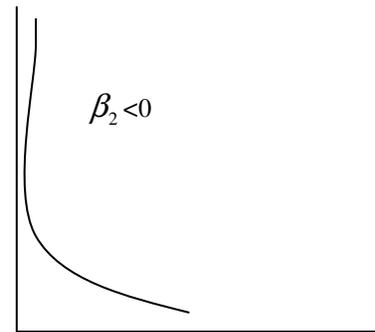
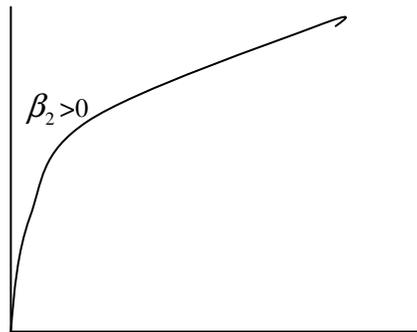
$$\ln \frac{Y_i}{1-Y_i} = \beta_1 + \beta_2 X_i$$

This model is known as the **logit model**, which we will discuss in the chapter on qualitative dependent variables.

6.12 (a)



(b)



- 6.13** (a) For every tenth of a unit increase (0.10) in the Gini coefficient, we would expect to see a 3.32 unit increase in a country's sociopolitical instability index. Therefore, as the Gini coefficient gets higher, or a country's income inequality gets larger, a country becomes less sociopolitically stable.
- (b) To see this difference, simply assess what happens if the Gini coefficient increases by 0.3. So,  $33.2 (0.3) = 9.96$ , indicating an increase of 9.96 in the SPI.
- (c) Using the standard  $t$  test,  $t = \frac{33.2}{11.8} = 2.8136$  for testing the null hypothesis that the slope coefficient is 0. For 38 degrees of freedom, the critical value from the table in Appendix D is somewhere between 2.021 and 2.042 (using a two-sided test), so the estimated slope is statistically significant at the 5% level.
- (d) Based on the regression results, we can conclude that there is a positive relationship between higher income inequality and greater political instability, although we cannot make a causal statement about the relationship.

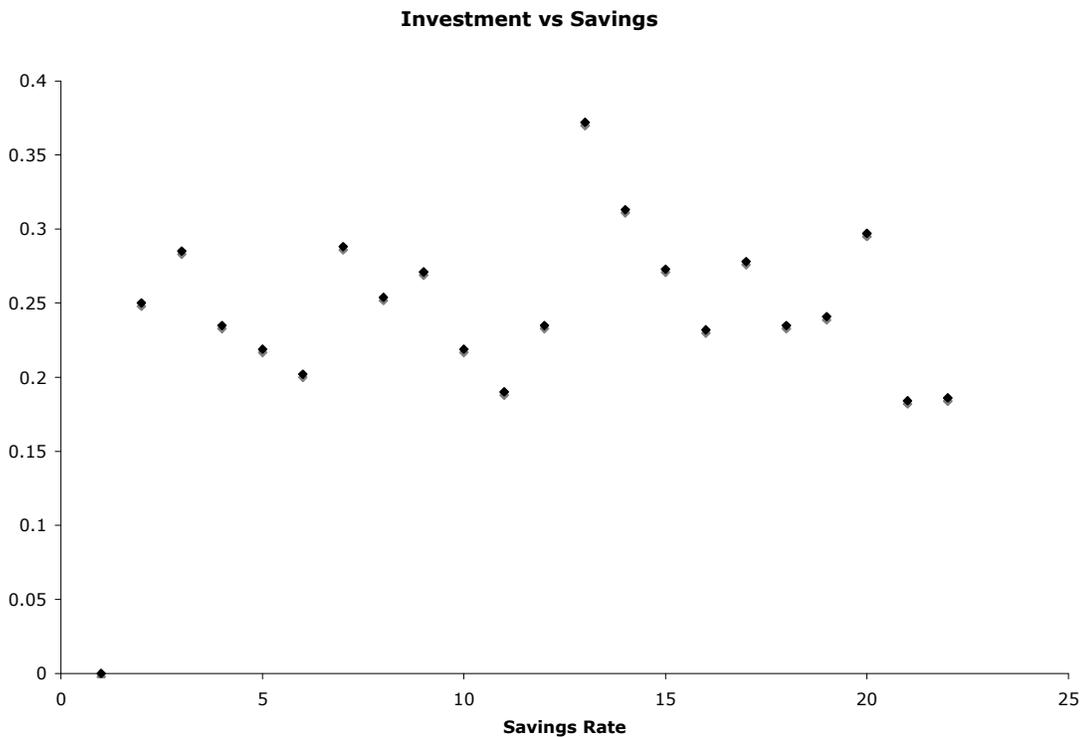
**Empirical Exercises**

$$6.14 \quad \frac{100}{100 - Y_i} = 2.0675 + 16.2662 \left( \frac{1}{X_i} \right)$$

$$se = (0.1596) (1.3232) \quad r^2 = 0.9497$$

As X increases indefinitely,  $\left( \frac{100}{100 - Y} \right)$  approaches the limiting value of 2.0675, which is to say that Y approaches the limiting value of about 51.6.

**6.15 (a)**



(b) Based on the scatterplot, there doesn't seem to be a very strong relationship between Savings and Investment. It actually seems like the same general level of investment occurs regardless of how much is being saved in each country. Therefore, neither the linear or log-linear models are likely to fit very well, although the existence of the observation in the lower lefthand corner of the plot may change these results.

(c) The regression results are as follows:

$$Invrate = 0.0435 + 0.8468 Savrate$$

$$se = (0.0176) (0.0693)$$

$$t = (2.4685) (12.222) \quad r^2 = 0.8872$$

$$\begin{aligned} \ln \text{Invrate} &= -0.02159 + 0.8288 \ln \text{Savrate} \\ \text{se} &= (0.0986)(0.0699) \\ t &= (-2.1901)(11.865) \quad r^2 = 0.8811 \end{aligned}$$

- (d) In the linear model, the slope coefficient can be interpreted as: If the savings rate increases by 0.1 (relative to GDP), the increase in investment expenditure (relative to GDP) will be about 0.0847, on average. Therefore, investment rates increase less than savings rates. For the log-linear model, a one percent increase in the Savings Rate generally corresponds to a 0.829 percent increase in the rate of Investment.
- (e) The intercept in the linear model suggests that, when the savings rate is 0 (no savings), a country's investment rate still exists, although it is small. This doesn't have much practical significance, though, since we don't see countries with 0 savings. The intercept in the log-linear model is negative, indicating that a 0% increase in savings should correspond to a drop in the percent of investment.
- (f) We cannot directly compare the  $r^2$  coefficients because the dependent variables are not the same.
- (g) For the linear model, the elasticity is not apparent. The log-linear model, however, already contains the results relative to the elasticities of the variables. To create the elasticity, we need to calculate the following:
- $$\frac{\partial \overline{\text{Invrate}}}{\partial \overline{\text{Savrate}}} \frac{\overline{\text{Savrate}}}{\overline{\text{Invrate}}} = 0.8468 \frac{\overline{\text{Savrate}}}{\overline{\text{Invrate}}}$$
- where the bar over the variables denotes their average values over the sample data.
- (h) Each model has its own usefulness, so it depends on the context or goal of the researcher.

## 6.16 (a)

Model	Slope estimate	se	t	r <sup>2</sup>
Linear	0.173	0.0058	29.666	0.3671
Log-linear	0.579	0.0187	30.958	0.3872
Log-lin	0.005	0.00002	26.808	0.3215
Lin-log	32.508	0.2586	125.720	0.9124
Reciprocal	-1652.22	56.8536	-29.061	0.3576
Log reciprocal	-50.498	1.6320	-30.942	0.3869

(b) We cannot compare the  $r^2$  values directly, but it does seem that the lin-log model has the best results. This would indicate that food expenditures are highly related to the elasticity (or percentage change) in total expenditures. This makes sense in that we would not expect to see a strong direct relationship between the two since food expenditures would probably not change much, given that they are a necessity. As total expenditures increases, so does food expenditures, but at a slower rate (as in a learning curve).

To obtain the growth rate of expenditure on durable goods, we can fit the log-lin model, whose results are as follows:

$$\begin{aligned} \ln \text{Expdur}_t &= 6.8950 + 0.0140 t \\ se &= (0.0082) (0.0009) \qquad r^2 = 0.9492 \end{aligned}$$

As this regression shows, over the sample period, the (quarterly) rate of growth in the durable goods expenditure was about 1.5 %. Both the estimated coefficients are individually statistically significant as the  $p$  values are extremely low. It would not make much sense to run a double log model here, such as:

$$\ln \text{Expdur}_t = \beta_1 + \beta_2 \ln \text{time} + u_t$$

Since the slope coefficient in this model is the elasticity coefficient, what is the meaning of the statement that as time increases by one percent, on average, expenditure on durable goods goes up by  $\beta_2$  percent?

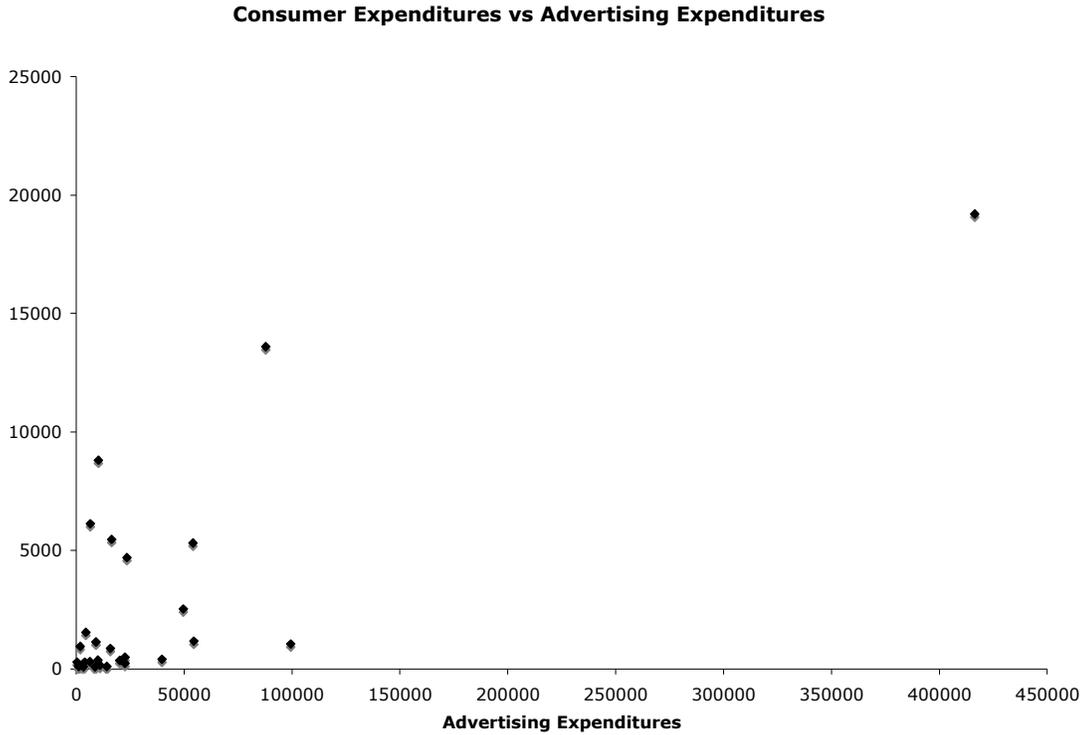
**6.18** The corresponding results for the non-durable goods sector are:

$$\begin{aligned} \ln \text{Expndur}_t &= 7.6257 + 0.0098 t \\ se &= (0.0021) (0.00023) \qquad r^2 = 0.9931 \end{aligned}$$

From these results it can be seen that over the sample period the (quarterly) rate of growth of expenditure on non-durables was about 0.98 percent. Comparing the results of the regressions in Problems 6.17 and 6.18, it seems that over the period 2003:01 to 2006:03, expenditure on durable goods increased at a much

faster rate than that on the non-durable goods. This may not be surprising in view of one of the longest economic expansions in the US history.

**6.19** (a) The scattergram of total consumer expenditure and advertising expenditure is as follows:



(b) Although the relationship between the two variables seems to be positive, it is not clear which particular curve will fit the data. In the following table we give regression results based on a few models.

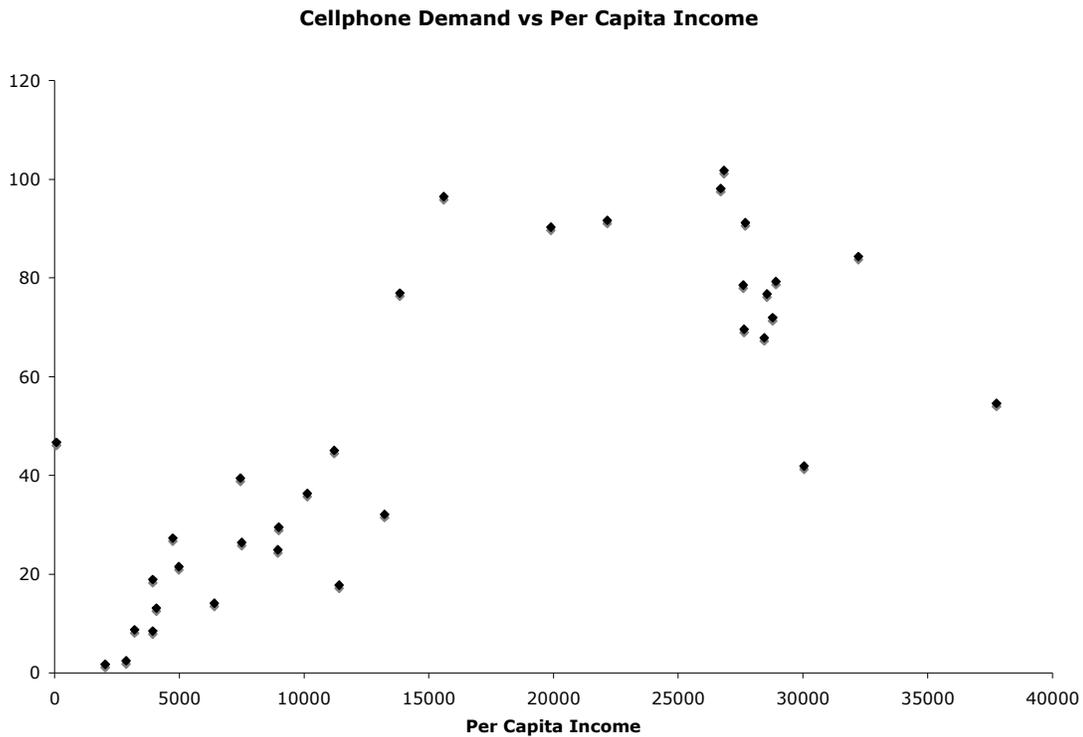
Model	Intercept	Slope	$r^2$
Linear	1057.361 (1.774)	0.0446 (6.283)	0.5938
Lin-log	-12585.01 (-2.872)	1626.643 (3.516)	0.3140
Reciprocal	3077.256 (3.344)	-1642108 (-1.143)	0.0461
Log-linear	0.9864 (0.628)	0.6038 (3.642)	0.3294
Log-lin	6.262 (21.354)	0.00001 (3.008)	0.2510
Log-recip	6.852 (20.951)	-797.845 (-1.563)	0.0829

*Note:* Figures in the parentheses are the estimated  $t$  values.  
 In each regression the regressand is total expenditure and the regressor is advertising expenditure.

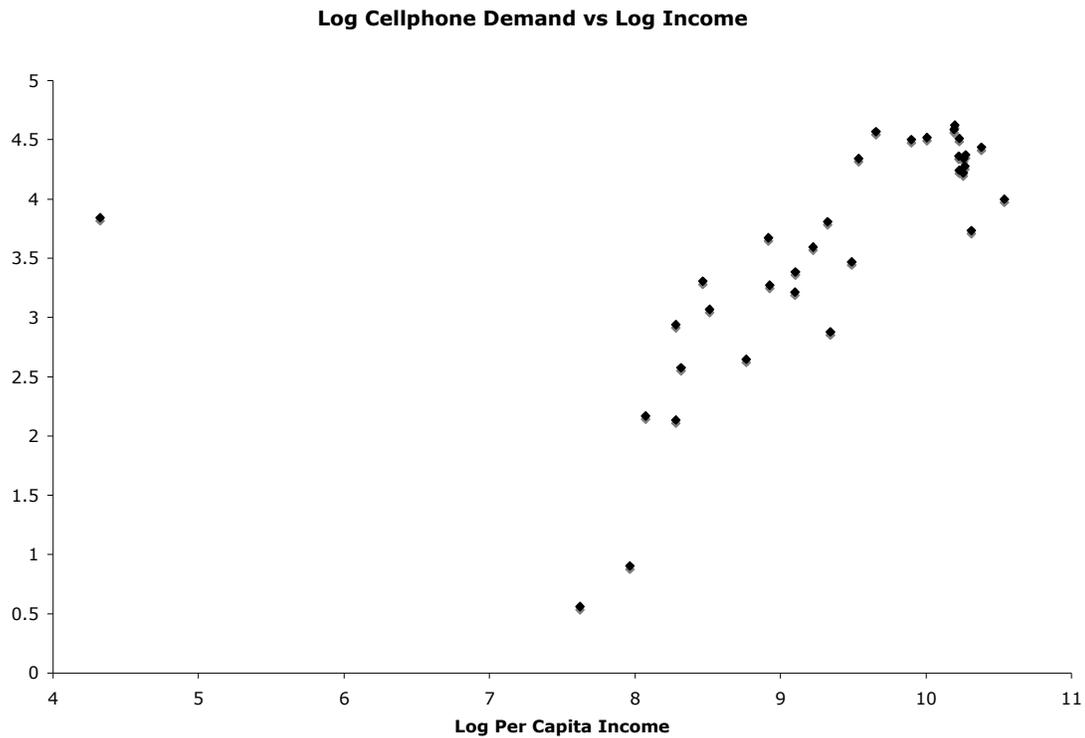
It is left to the reader to compare the various models. Note that the  $r^2$  values of the first two models are comparable, since the regressand is the same in the two models. Similarly, the  $r^2$ s of the last three models are comparable (Why?)

(c) Assessing the ratio of the variables, it seems there are a few unusually high values. The average ratio is 0.0342, with a standard deviation of 0.0396. There are high values of 0.0946, 0.1051, 0.0972, 0.1512, and 0.0924. These could definitely affect the regression results.

**6.20 (a)**



(b)



(c) The first graph seems to exhibit non-constant variance, whereas the second graph appears to be more constant. In the second graph, the effect of Bulgaria is even more extreme (the furthest left point), suggesting that it might be an outlier.

(d) Aside from the issue discussed in part (c), the second model would most likely give smaller standard errors, and therefore more consistent estimates.

(e) Double-log regression results are:

$$\ln \text{Cellphone}_i = -0.9374 + 0.4864 \ln \text{PCIncome}_i$$

$$t = (-0.830) \quad (4.018) \quad r^2 = 0.3353$$

The slope of  $\ln \text{PCIncome}$  indicates the elasticity, so for a one percent change in  $\text{PCIncome}$ , we would expect to see about a 0.486 percent increase in Cellphone demand.

(f) Yes, it is statistically significant with a *p value* of about 0.000.

(g) For the linear model, the elasticity is not directly apparent. The log-linear model, however, already contains the results relative to the elasticities of the variables. To create the elasticity, we need to calculate the following:

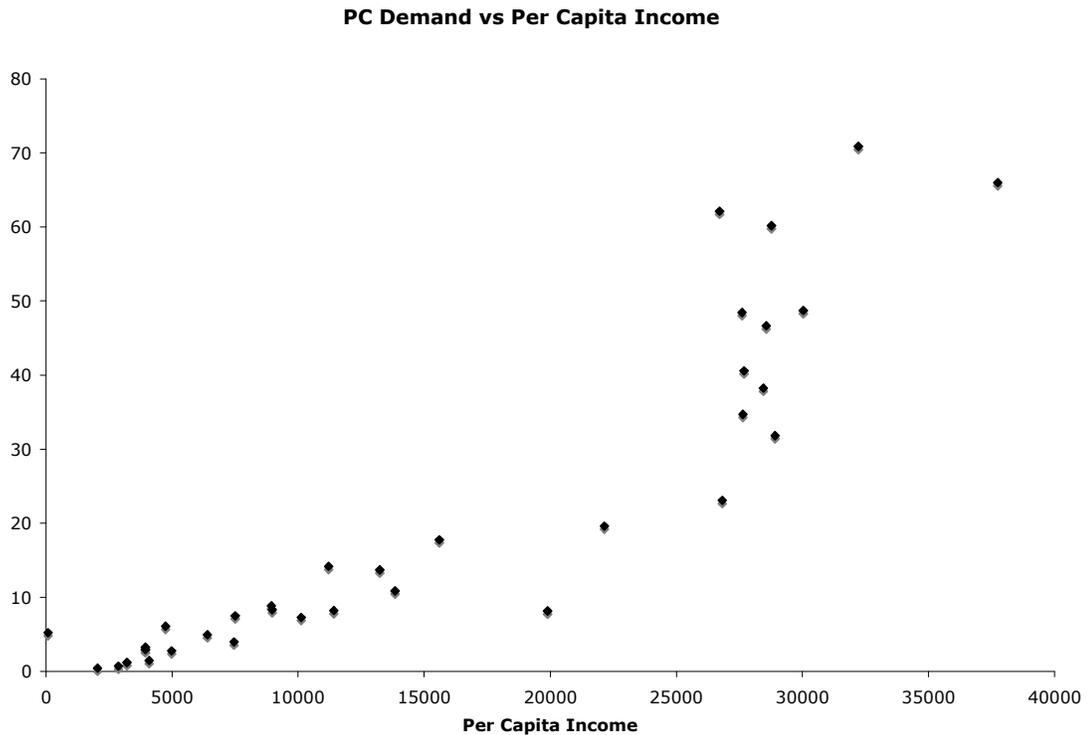
$$\frac{\frac{\partial \overline{Cellphone}}{\partial \overline{PCIncome}}}{\overline{Cellphone}} = 0.0022 \frac{\overline{PCIncome}}{\overline{Cellphone}}$$

where the bar over the variables denotes their average values over the sample data. Therefore, we need to know the average values of the variables to compute this. They are: average PCIncome = 15819.865 and average Cellphone demand = 49.574. The calculated elasticity is then

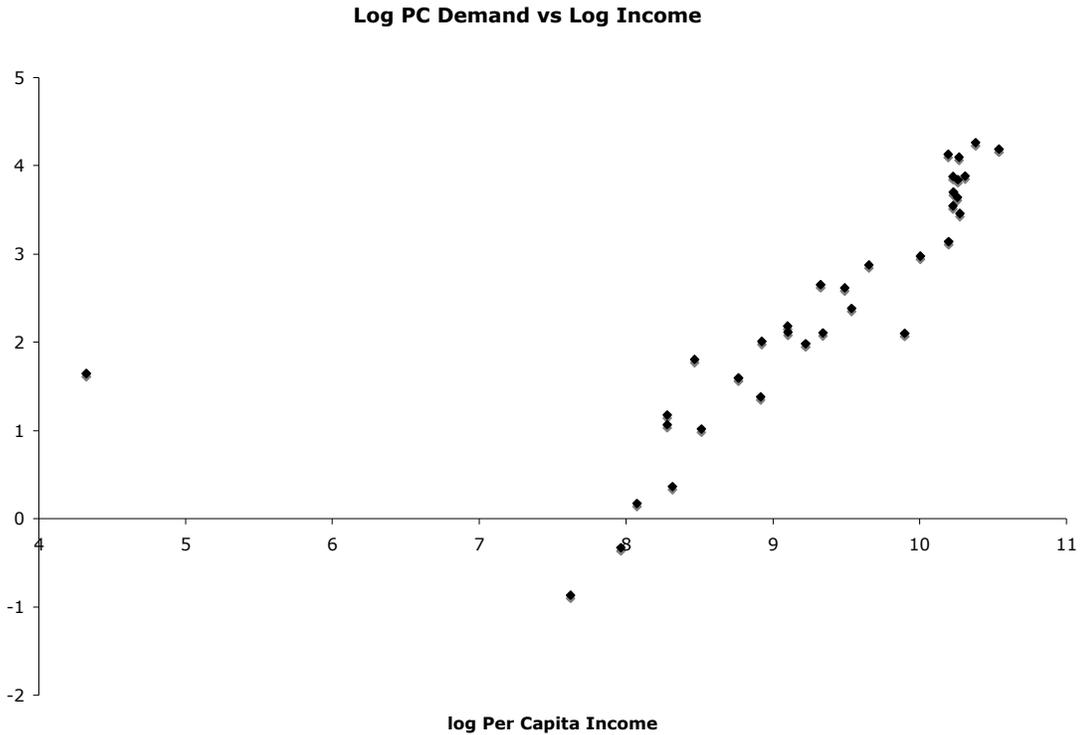
$$elasticity = 0.0022 \frac{15819.865}{49.574} = 0.7021$$

(h) There is a difference... the better choice would probably be to use the double-log model instead of backing out of the linear model.

6.21 (a)



(b)



(c) The first graph seems to exhibit non-constant variance, whereas the second graph appears to be more constant. In the second graph, the effect of Bulgaria is even more extreme (the furthest left point), suggesting that it might be an outlier.

(d) Aside from the issue discussed in part (c), the second model would most likely give smaller standard errors, and therefore more consistent estimates.

(e) Double-log regression results are:

$$\ln PC_i = -5.2712 + 0.8269 \ln PCIncome_i$$

$$t = (-4.181) \quad (6.116) \quad r^2 = 0.5390$$

The slope of  $\ln PCIncome$  indicates the elasticity, so for a one percent change in  $PCIncome$ , we would expect to see about a 0.827 percent increase in  $PC$  demand. It would seem that, in general,  $PC$  demand is more quickly affected by income changes than cellphone demand. It is possible that cellphones are now so ubiquitous that the demand is relatively inelastic.

(f) Yes, it is statistically significant with a  $p$  value of about 0.000.

(g) For the linear model, the elasticity is not directly apparent. The log-linear model, however, already contains the results relative to the elasticities of the variables. To create the elasticity, we need to calculate the following:

$$\frac{\frac{\partial PC_s}{\partial PCIncome}}{\frac{\overline{PCIncome}}{\overline{PC_s}}} = 0.0018 \frac{\overline{PCIncome}}{\overline{PC_s}}$$

where the bar over the variables denotes their average values over the sample data. Therefore, we need to know the average values of the variables to compute this. They are: average PCIncome = 15819.865 and average PC demand = 21.429. The calculated elasticity is then

$$elasticity = 0.0018 \frac{15819.865}{21.429} = 1.329$$

- (i) There is a difference... the better choice would probably be to use the double-log model instead of backing out of the linear model.

**6.22** (a) Linear regression results are:

$$\begin{aligned} \text{Cellphone}_i &= 29.2342 + 0.9492 \text{PC}_s_i & r^2 &= 0.4168 \\ t &= (4.851) \quad (4.782) \end{aligned}$$

- (b) Yes, it is statistically significant with a *p* value of about 0.000.
- (c) Running the opposite regression would give the same  $r^2$  results and an inverted slope value. Basically, the results are identical.
- (d) It would really depend on which variable is considered to be a given and which is considered to be dependent. The choice would be based solely on the goal of the researcher.