THE TWO-VARIABLE LINEAR REGRESSION MODEL

Herman J. Bierens

Pennsylvania State University

April 30, 2012

1. *Introduction*

Suppose you are an economics or business major in a college close to the beach in the southern part of the US, for example southern California[1], where the weather is almost always nice the whole year around. In order to support yourself through college, you have started your own (weekend) business: an ice cream parlor on the beach. You have experienced that on hot weekends you usually sell more ice cream than on cold weekends. Also, you have recorded the average temperature and the sales of ice cream during eight weekends. Let Y_j be the sales of ice cream on weekend j, measured in \$100, and let X_j be the average temperature on weekend j, measured in units of 10 degrees Fahrenheit:

Table 1: *Ice cream data*

Sales (unit = \$100)	Temperature (unit = 10 degrees)
$Y_1 = 8$	$X_1 = 5$
$Y_2 = 10$	$X_2 = 7$
$Y_3 = 8$	$X_3 = 6$
$Y_4 = 13$	$X_4 = 8$
$Y_5 = 15$	$X_5 = 10$
$Y_6 = 14$	$X_6 = 9$
$Y_7 = 11$	$X_7 = 7$
$Y_8 = 9$	$X_8 = 8$

You want to use this information to forecast next weekend's sales of ice cream, given a good forecast of next weekend's temperature. Such a forecast of the sales will enable you to

---

[1] These lecture notes are based on lecture notes that I wrote while teaching at the University of California, San Diego, in the winter of 1987.

reduce your cost by adjusting your purchase of ice cream to the expected demand, because the ice cream you don't sell has to be thrown away.

Let your forecasting scheme be

$$\hat{Y} = \hat{\alpha} + \hat{\beta}.X,$$

i.e., given the temperature of X times 10 degrees and given the values of $\hat{\alpha}$ and $\hat{\beta}$, $\hat{Y}$ times \$100 will be your forecast of the sales of ice cream. This forecasting scheme together with the points (X_j, Y_j), $j = 1,2,...,8,$ is plotted in Figure 1:

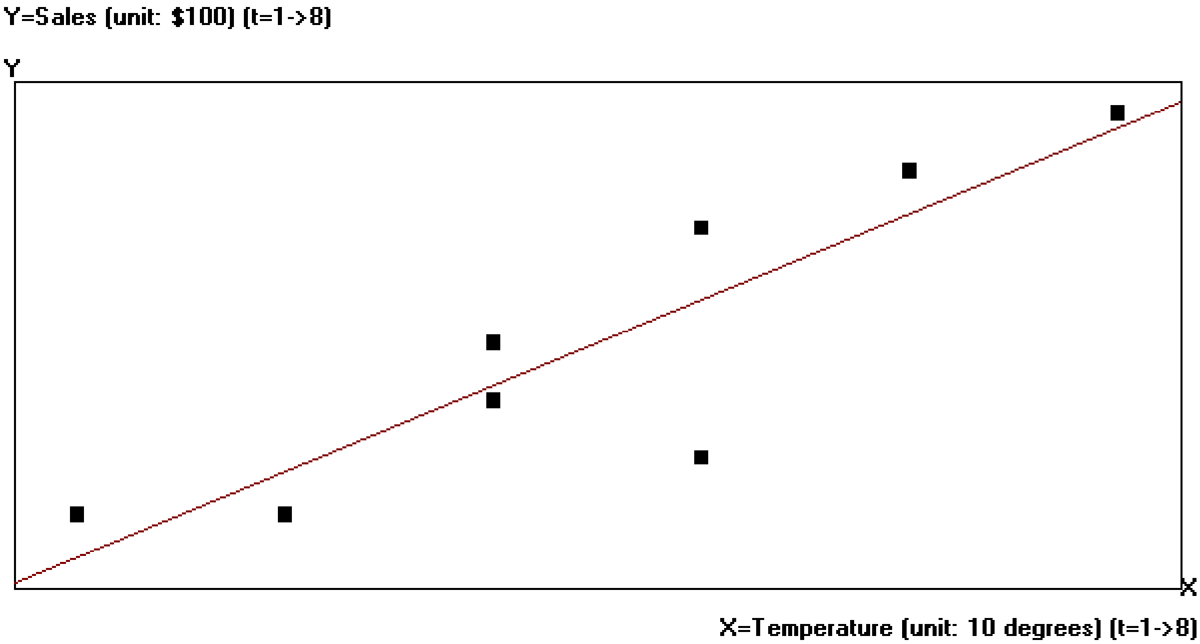

Figure 1 Scatter plot of (X_j, Y_j), $j = 1,2,...,8,$ together with the line $\hat{Y} = \hat{\alpha} + \hat{\beta}.X.$

The best values for $\hat{\alpha}$ and $\hat{\beta}$ are those for which the forecast error (= actual sales minus forecasted sales) is minimal. However, you do not know yet the actual sales in the next weekend, but you do know the actual sales in the eight weekends for which you have recorded your sales and the corresponding temperature. So what you could do is to forecast the sales of ice cream on each of these eight weekends and to determine $\hat{\alpha}$ and $\hat{\beta}$ such that the forecast errors are minimal. Because forecast errors can be positive and negative, as can be seen from Figure 1, the sum of the forecast errors is not a good measure of the performance of your forecasting scheme, because large positive errors can be offset by large negative errors. Therefore, use the sum of squared errors as your measure of the accuracy of your forecasts:

$$Q(\hat{\alpha},\hat{\beta}) = \sum_{j=1}^{n} (Y_j - \hat{Y}_j)^2 = \sum_{j=1}^{n} (Y_j - \hat{\alpha} - \hat{\beta}X_j)^2,$$

where n is the sample size ($n = 8$ in our example), and minimize $Q(\hat{\alpha},\hat{\beta})$ to $\hat{\alpha}$ and $\hat{\beta}$. It can be shown (see the Appendix) that $Q(\hat{\alpha},\hat{\beta})$ is minimal for

$$\hat{\beta} = \frac{\sum_{j=1}^{n}(X_j - \bar{X})(Y_j - \bar{Y})}{\sum_{j=1}^{n}(X_j - \bar{X})^2} = \frac{\sum_{j=1}^{n}(X_j - \bar{X})Y_j}{\sum_{j=1}^{n}(X_j - \bar{X})^2}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}, \text{ where } \bar{X} = (1/n)\sum_{j=1}^{n}X_j \text{ and } \bar{Y} = (1/n)\sum_{j=1}^{n}Y_j. \tag{1}$$

In the ice cream parlor case we have

$$n = 8,\ \bar{X} = 7.5,\ \bar{Y} = 11,\ \sum_{j=1}^{n}X_j^2 = 468,\ \sum_{j=1}^{n}X_jY_j = 687,$$

$$\sum_{j=1}^{n}(X_j - \bar{X})(Y_j - \bar{Y}) = \sum_{j=1}^{n}X_jY_j - n.\bar{X}.\bar{Y} = 27,$$

$$\sum_{j=1}^{n}(X_j - \bar{X})^2 = \sum_{j=1}^{n}X_j^2 - n.\bar{X}^2 = 18,$$

so that

$$\hat{\beta} = 1.5, \quad \hat{\alpha} = -0.25.$$

Thus, our best forecasting scheme is $\hat{Y} = -0.25 + 1.5X$. This is the straight line in Figure 1.

Now suppose that the forecast of next weekend's temperature is 75 degrees. Then $X = 7.5$, hence $\hat{Y} = -0.25 + 1.5\ (7.5) = 11$. Therefore, the best forecast of next weekend's sales is: $\hat{Y}\ x\ \$100 = \$1,100$.

2. *The two-variable linear regression model.*

In order to answer the question how good this forecast is, we have to make assumptions about the true relationship between the *dependent variable* Y_j and the *independent variable* X_j, (also called the *explanatory variable*). The true relationship we are going to assume is the two-variable linear regression model:

$$Y_j = \alpha + \beta.X_j + U_j, \quad j = 1,2,.....,n. \tag{2}$$

The U_j's are random error variables, called *error terms*, for which we assume:

Assumption I: *The U_j's are independent and identically distributed (i.i.d) random variables.*

Assumption II: *The mathematical expectation of U_j equals zero*: $E(U_j) = 0$ *for* $j = 1,2,...,n$.

Assumption III: *The variance* $\sigma^2 = \text{var}(U_j) = E[(U_j - E(U_j))^2] = E[U_j^2]$ *of the U_j's is constant and finite.*

Regarding the explanatory variables X_j we shall assume for the time being that

Assumption IV: *The independent variables X_j are non-random.*

This assumption is not strictly necessary, and is actually quite unrealistic in economics, but will be made for the sake of convenience, as it will ease the argument. Finally, we will assume that the errors are normally distributed:

Assumption V: *The errors U_j's are* $N(0, \sigma^2)$ *distributed.*

In particular, we shall need the latter assumption in order to say something about the reliability of the forecast. These assumptions will be relaxed later on.

3. *The properties of* $\hat{\alpha}$ *and* $\hat{\beta}$.

Although we have motivated model (2) by the need to forecast out-of-sample values of the dependent variables Y_j, a linear regression model is more often used for testing economic hypotheses. For example, let Y_j be the hourly wage of wage earner j in a random sample of size n of wage earners, and let X_j be a gender indicator, say $X_j = 1$ if person j is a female, and $X_j = 0$ if person j is a male. If you suspect gender discrimination in the workplace, you can test this suspicion by testing the null hypothesis that $\beta = 0$ (no gender discrimination) against one of three possible alternative hypotheses:

(a) $\beta \neq 0$: women are paid different hourly wages than men, either higher or lower;

(b) $\beta > 0$: women are paid higher hourly wages than men;

(c) $\beta < 0$: women are paid lower hourly wages than men.

The last hypothesis is usually what is meant by "gender discrimination." A test for the null hypothesis $\beta = 0$ against one of these alternative hypotheses can be based on the estimate $\hat{\beta}$ of β, provided that we know how $\hat{\beta}$ is related to β.

It will be shown below that $\hat{\alpha}$ and $\hat{\beta}$ are indeed reasonable approximations of α and β, respectively, possessing particular desirable properties.

In general an *estimator* of an unknown parameter is a function of the data that serves as an approximation of the parameter involved. It follows from (1) that $\hat{\alpha}$ and $\hat{\beta}$ are functions of the data, $(Y_1,X_1),\ldots,(Y_n,X_n)$. Because $\hat{\alpha}$ and $\hat{\beta}$ will be used as approximations of α and β, respectively, and were obtained by minimizing the squared errors, we will call $\hat{\alpha}$ and $\hat{\beta}$ the Ordinary[2] Least Squares (OLS) estimators of α and β, respectively.

3.1 *Unbiasedness*

The first property of $\hat{\alpha}$ and $\hat{\beta}$ is that they are *unbiased* estimators of α and β:

Proposition 1. *Under Assumptions II and IV the OLS estimators* $\hat{\alpha}$ *and* $\hat{\beta}$ *are unbiased*, *which means that* $E[\hat{\alpha}] = \alpha$ *and* $E[\hat{\beta}] = \beta$.

This result follows from the fact that we can write

$$\hat{\alpha} = \alpha + \sum_{j=1}^{n}\left(\frac{1}{n} - \frac{\bar{X}(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right).U_j, \quad \hat{\beta} = \beta + \frac{\sum_{j=1}^{n}(X_j-\bar{X})U_j}{\sum_{i=1}^{n}(X_i-\bar{X})^2}. \tag{3}$$

See the Appendix.

[2] The estimators $\hat{\alpha}$ and $\hat{\beta}$ are called "Ordinary" least squares estimators to distinguish them from "Nonlinear" least squares estimators.

3.2 *The variances of* $\hat{\alpha}$ *and* $\hat{\beta}$.

Our next issue concerns the variances of $\hat{\alpha}$ and $\hat{\beta}$. For deriving these variances the following two lemmas are convenient.

Lemma 1. *Let* $U_1, U_2,...,U_n$ *be independent random variables with zero mathematical expectation (thus* $E(U_j) = 0$*) and variance* σ^2. *(Thus* $E[(U_j - E(U_j))^2] = E(U_j^2) = \sigma^2$*). Let* $v_1, v_2,...,v_n$ *and* $w_1, w_2,...,w_n$ *be given constants. Then* $E[(\sum_{j=1}^n v_j U_j)(\sum_{j=1}^n w_j U_j)] = \sigma^2 \sum_{j=1}^n v_j w_j$.

Proof. See the Appendix.

Note that if we choose $v_j = w_j$ for $j = 1,2,...,n$ in Lemma 1 then it reads:

Lemma 2. *Let* $U_1, U_2,...,U_n$ *be independent random variables with zero mathematical expectation and variance* σ^2. *Let* $w_1, w_2,...,w_n$ *be given constants. Then* $E[(\sum_{j=1}^n w_j U_j)^2] = \sigma^2 \sum_{j=1}^n w_j^2$.

Using (3) and Lemmas 1 and 2 it can be shown that

Proposition 2. *Under the assumptions I - IV,*

$$\mathrm{var}(\hat{\alpha}) = \frac{\sigma^2 \sum_{j=1}^n X_j^2}{n \sum_{j=1}^n (X_j - \bar{X})^2} = \sigma_{\hat{\alpha}}^2 \text{, say, } \mathrm{var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{j=1}^n (X_j - \bar{X})^2} = \sigma_{\hat{\beta}}^2 \text{, say, and}$$

$$\mathrm{cov}(\hat{\alpha}, \hat{\beta}) = \frac{-\sigma^2 \bar{X}}{\sum_{j=1}^n (X_j - \bar{X})^2}. \tag{4}$$

Proof. See the Appendix

3.3 *Normality of* $\hat{\alpha}$ *and* $\hat{\beta}$.

If we also assume normality of the error terms U_j then $\hat{\alpha}$ and $\hat{\beta}$ are also normally distributed. This result follows from the following lemma.

Lemma 3. *Let Z_1, Z_2,...Z_m be independent $N(\mu,\sigma^2)$ distributed random variables and let w_1,..,w_m be constants. Then* $\sum_{j=1}^{m} w_j Z_j$ *is distributed* $N[(\sum_{j=1}^{m} w_j)\mu, (\sum_{j=1}^{m} w_j^2)\sigma^2]$.

The proof of this lemma requires advanced probability theory and is therefore omitted.

It follows now straightforwardly from Proposition 2, Lemma 3, and (3) that:

Proposition 3. *Under the assumptions I - V,*

$$\hat{\alpha} - \alpha \sim N\left[0, \frac{\sigma^2 \sum_{j=1}^{n} X_j^2}{n \sum_{j=1}^{n} (X_j - \bar{X})^2}\right], \quad \hat{\beta} - \beta \sim N\left[0, \frac{\sigma^2}{\sum_{j=1}^{n} (X_j - \bar{X})^2}\right], \tag{5}$$

where " ~ " *is the symbol for* "*is distributed as.*"

Moreover, applying Lemma 3 again for $m = 1$ it follows from (5) (Exercise: Why?) that

Proposition 4. *Under the assumptions I - V,*

$$\frac{(\hat{\alpha} - \alpha)\sqrt{n \sum_{j=1}^{n} (X_j - \bar{X})^2}}{\sigma . \sqrt{\sum_{j=1}^{n} X_j^2}} \sim N[0,1], \quad \frac{(\hat{\beta} - \beta)\sqrt{\sum_{j=1}^{n} (X_j - \bar{X})^2}}{\sigma} \sim N[0,1]. \tag{6}$$

These results play a key-role in testing hypotheses about α and β. The only problem that prevents us from using these results for testing is that σ is unknown. This problem will be addressed in the next section.

4. *How to estimate the error variance* σ^2 ?

If α and β were known then we could estimate σ^2 by

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} (Y_j - \alpha - \beta . X_j)^2 = \frac{1}{n} \sum_{j=1}^{n} U_j^2 . \tag{7}$$

However, α and β are not known, but we do have OLS estimators of α and β. This suggests to replace α and β in (7) by their OLS estimators:

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \hat{\alpha} - \hat{\beta}.X_j)^2 = \frac{1}{n}\sum_{j=1}^{n}\hat{U}_j^2, \tag{8}$$

where

$$\hat{U}_j = Y_j - \hat{\alpha} - \hat{\beta}.X_j \tag{9}$$

is called the regression *residual.* However, the estimator (8) is biased, due to the fact that

Proposition 5. *Under the assumptions I - V,* $E[\sum_{j=1}^{n}\hat{U}_j^2] = (n - 2)\sigma^2$.

Proof: See the Appendix.

This result suggests to use

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{j=1}^{n}\hat{U}_j^2 \tag{10}$$

as an estimator of σ^2 instead of (8), because then by Proposition 5, $\hat{\sigma}^2$ is an unbiased estimator of σ^2:

$$E[\hat{\sigma}^2] = \sigma^2. \tag{11}$$

The sum $\sum_{j=1}^{n}\hat{U}_j^2$ is called the Sum of Squares Residuals, shortly *SSR*, or also called the Residual Sum of Squares (RSS), and $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ is called the Standard Error of the Residuals, shortly *SER*. Thus,

$$SSR = \sum_{j=1}^{n}\hat{U}_j^2, \qquad SER = \sqrt{\frac{\sum_{j=1}^{n}\hat{U}_j^2}{n-2}} = \sqrt{\frac{SSR}{n-2}} \ (= \hat{\sigma}). \tag{12}$$

Finally, note that the sum of squared residuals can be computed as follows:

$$SSR = \sum_{j=1}^{n}(Y_j - \bar{Y})^2 - \hat{\beta}^2\sum_{j=1}^{n}(X_j - \bar{X})^2. \tag{13}$$

See the Appendix.

5. *Standard errors, t-values and p-values of the OLS estimators*

The variances of $\hat{\alpha}$ and $\hat{\beta}$ can now be estimated by replacing σ^2 in (4) by $\hat{\sigma}^2$:

$$\textit{Estimated}\ \mathrm{var}(\hat{\alpha}) = \frac{\hat{\sigma}^2 \sum_{j=1}^{n} X_j^2}{n\sum_{j=1}^{n}(X_j - \bar{X})^2} = \hat{\sigma}_{\hat{\alpha}}^2,\ \textit{say},$$

$$\textit{Estimated}\ \mathrm{var}(\hat{\beta}) = \frac{\hat{\sigma}^2}{\sum_{j=1}^{n}(X_j - \bar{X})^2} = \hat{\sigma}_{\hat{\beta}}^2\ ,\ \textit{say}. \tag{14}$$

Then $\hat{\sigma}_{\hat{\alpha}} = \sqrt{\hat{\sigma}_{\hat{\alpha}}^2}$ is called the standard error of $\hat{\alpha}$, also denoted by $SE(\hat{\alpha})$, and $\hat{\sigma}_{\hat{\beta}} = \sqrt{\hat{\sigma}_{\hat{\beta}}^2}$ is called the standard error of $\hat{\beta}$, also denoted by $SE(\hat{\beta})$.

If we replace σ in Proposition 4 by the SER, $\hat{\sigma}$, the standard normality results involved change:

Proposition 6. *Under the assumptions I - V,*

$$\frac{\hat{\alpha}-\alpha}{\hat{\sigma}_{\hat{\alpha}}} = \frac{(\hat{\alpha} - \alpha)\sqrt{n\sum_{j=1}^{n}(X_j - \bar{X})^2}}{\hat{\sigma}.\sqrt{\sum_{j=1}^{n}X_j^2}} \sim t_{n-2},\quad \frac{\hat{\beta}-\beta}{\hat{\sigma}_{\hat{\beta}}} = \frac{(\hat{\beta} - \beta)\sqrt{\sum_{j=1}^{n}(X_j - \bar{X})^2}}{\hat{\sigma}} \sim t_{n-2}. \tag{15}$$

The proof of Proposition 6 is based on the fact that under these assumptions, SSR/σ^2 is distributed χ^2_{n-2} and is independent of $\hat{\alpha}$ and $\hat{\beta}$, but the proof involved requires advanced probability theory and is therefore omitted.

Because for large degrees of freedom the *t* distribution is approximately equal to the standard normal distribution, and due to the central limit theorem, Proposition 4 holds if *n* is large and the errors are not normally distributed, we also have:

Proposition 7. *If the sample size n is large then under the assumptions I - IV we have approximately,*

$$\frac{\hat{\alpha}-\alpha}{\hat{\sigma}_{\hat{\alpha}}} = \frac{(\hat{\alpha} - \alpha)\sqrt{n\sum_{j=1}^{n}(X_j - \bar{X})^2}}{\hat{\sigma}.\sqrt{\sum_{j=1}^{n}X_j^2}} \sim N(0,1),\quad \frac{\hat{\beta}-\beta}{\hat{\sigma}_{\hat{\beta}}} = \frac{(\hat{\beta} - \beta)\sqrt{\sum_{j=1}^{n}(X_j - \bar{X})^2}}{\hat{\sigma}} \sim N(0,1). \tag{16}$$

The results in Proposition 6 now enable us to test hypotheses about α and β. In particular the null hypothesis that $\beta = 0$ is of importance, because this hypothesis implies that *X* has no effect on *Y*. The test statistic for testing this hypothesis is the t-value (or t-statistic) of $\hat{\beta}$:

$$\hat{t}_{\hat{\beta}} \ (= \ t\text{-}value \ \ of \ \hat{\beta}) \ \stackrel{def.}{=} \ \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} \ = \ \frac{\hat{\beta}\sqrt{\sum_{j=1}^{n}(X_j - \bar{X})^2}}{\hat{\sigma}} \ \sim \ t_{n-2} \ \ if \ \beta \ = \ 0. \tag{17}$$

If $\beta > 0$ and $n \to \infty$ then the t-value of $\hat{\beta}$ converges in probability to $+\infty$, and if $\beta < 0$ and $n \to \infty$ then the t-value of $\hat{\beta}$ converges in probability to $-\infty$. Moreover, if the sample size *n* is large then by Proposition 7 we may use the standard normal distribution instead of the t distribution to find critical values of the test.

Similarly,

$$\hat{t}_{\hat{\alpha}} \ (= \ t\text{-}value \ \ of \ \hat{\alpha}) \ \stackrel{def.}{=} \ \frac{\hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}} \ \sim \ t_{n-2} \ \ if \ \alpha \ = \ 0. \tag{18}$$

However, the hypothesis $\alpha = 0$ is often of no interest.

In the ice cream example,

$$\sum_{j=1}^{n}(X_j \ - \ \bar{X})^2 \ = \ 18 \ \Rightarrow \ \sqrt{\sum_{j=1}^{n}(X_j \ - \ \bar{X})^2} \ = \ \sqrt{18} \ \approx \ 4.24264,$$

$$\sum_{j=1}^{n}(Y_j \ - \ \bar{Y})^2 \ = \ \sum_{j=1}^{n}Y_j^2 \ - \ n.\bar{Y}^2 \ = \ 1020 \ - \ 8\times 11^2 \ = \ 52$$

and by (13),

$$\hat{\sigma}^2 \ = \ \frac{1}{n-2}\sum_{j=1}^{n}\hat{U}_j^2 \ = \ \frac{1}{n-2}\sum_{j=1}^{n}(Y_j \ - \ \bar{Y})^2 \ - \ \hat{\beta}^2 \ \frac{1}{n-2}\sum_{j=1}^{n}(X_j - \bar{X})^2$$

$$= \ \frac{52 \ - \ (1.5)^2.18}{8-2} \ = \ \frac{11.5}{6} \ \approx \ 1.916667 \ \Rightarrow \ \hat{\sigma} \ \approx \ 1.384437$$

Hence,

$$\hat{t}_{\hat{\beta}} \ = \ \frac{\hat{\beta}\sqrt{\sum_{j=1}^{n}(X_j - \bar{X})^2}}{\hat{\sigma}} \ = \ \frac{1.5\times 4.24264}{1.384437} \ \approx \ 4.597 \tag{19}$$

Assuming that the conditions of Proposition 6 hold, the null hypothesis $H_0\!: \ \beta \ = \ 0$ can be tested against the alternative hypothesis H_1: $\beta \neq 0$ using the two-sided t-test at say the 5% significance level, as follows. Under the null hypothesis, (19) is a random drawing from the *t* distribution with $n-2 = 6$ degrees of freedom. Look up in the table of the *t* distribution the value t_* such that for $T \sim t_6$, $P[|T| > t_*] = 0.05$. This value is $t_* = 2.447$. Then accept the null hypothesis if $-t_* = -2.447 \leq \hat{t}_{\hat{\beta}} \leq 2.447 = t_*$, and reject the null hypothesis in favor of the alternative hypothesis if $|\hat{t}_\beta| > t_* = 2.447$. Thus, in the ice cream example we reject the null hypothesis H_0: $\beta = 0$ because $|\hat{t}_\beta| = 4.597 > 2.447 = t_*$.

This test is illustrated in Figure 2 below. The curved line in Figure 2 is the density of the *t* distribution with 6 degrees of freedom. The grey areas are each 0.025, so that the total grey area is 0.05.

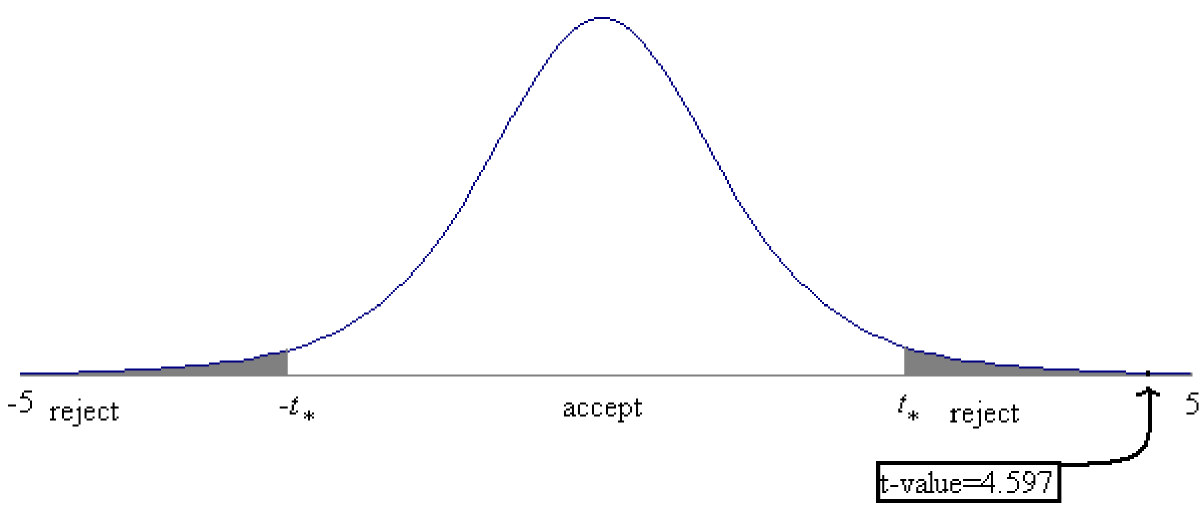

Figure 2 Two-sided t-test of H_0: $\beta = 0$ against the alternative hypothesis H_1: $\beta \neq 0$.

The null hypothesis H_0: $\beta = 0$ can be tested against the alternative hypothesis H_1: $\beta > 0$ at the 5% significance level by the right-sided t-test. Now look up in the table of the *t* distribution the value t_* such that for $T \sim t_6$, $P[T > t_*] = 0.05$. This value corresponds to the critical value of the two-sided t-test at the 10% significance level: $t_* = 1.943$. Then accept the null hypothesis if $\hat{t}_{\hat{\beta}} \leq t_* = 1.943$, and reject the null hypothesis in favor of the alternative hypothesis if $\hat{t}_\beta > t_* = 1.943$. Thus, in the ice cream case we reject the null hypothesis

$H_0: \beta = 0$ in favor of the alternative hypothesis $H_1: \beta > 0$.

This right-sided t-test is illustrated in Figure 3 below. Again, the curved line in Figure 3 is the density of the *t* distribution with 6 degrees of freedom, and the grey area is 0.05.

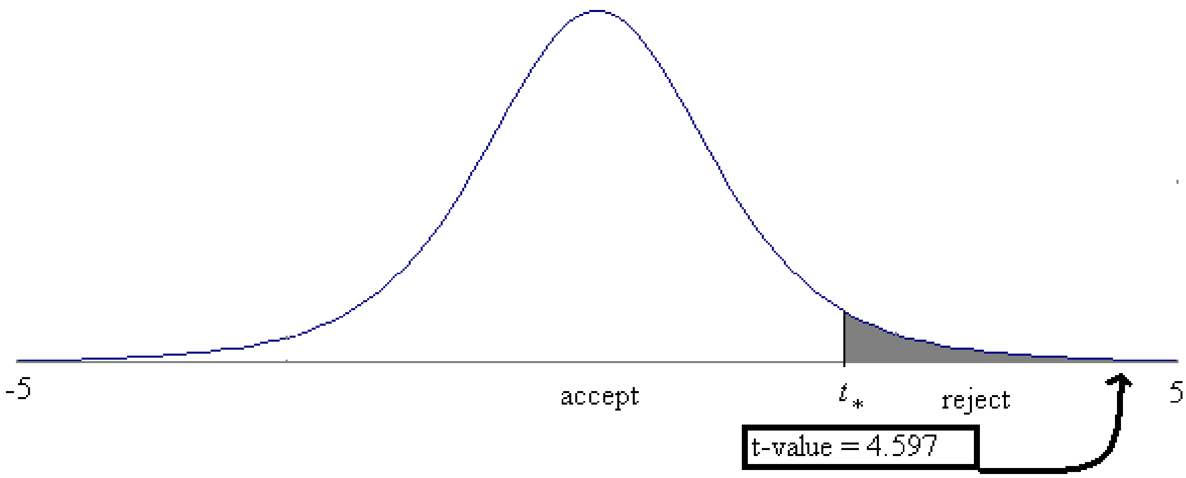

Figure 3 Right-sided t-test of $H_0: \beta = 0$ against the alternative hypothesis $H_1: \beta > 0$.

If the sample size *n* is large, so that $\hat{t}_{\hat{\beta}} \sim N(0,1)$ if $\beta = 0$, then an alternative way of testing the null hypothesis $\beta = 0$ against the alternative hypothesis $\beta \neq 0$ is to use the (two-sided) p-value:

$$\hat{p}_{\hat{\beta}} \ (= p\text{-}value \ of \ \hat{\beta}) \overset{def.}{=} P[|U| > |\hat{t}_{\hat{\beta}}|], \ where \ U \sim N(0,1). \tag{20}$$

For example, if $\hat{p}_{\hat{\beta}} < 0.05$ we reject the null hypothesis $\beta = 0$ in favor of the alternative hypothesis $\beta \neq 0$ at the 5% significance level, and if $\hat{p}_{\hat{\beta}} \geq 0.05$ we accept the null hypothesis $\beta = 0$. The p-value for $\hat{\alpha}$ is defined and used similarly.

Although a t-value is a test statistics of the null hypothesis that the corresponding coefficient in the regression model is zero, it is quite easy to rebuild the t-value for testing other null hypotheses, as follows. Suppose you want to test the null hypothesis that $\beta = \beta_0$, where β_0 is a given number, for example $\beta_0 = 1$. Then

$$\frac{\hat{\beta}-\beta_0}{\hat{\sigma}_{\hat{\beta}}} = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} - \frac{\beta_0}{\hat{\sigma}_{\hat{\beta}}} = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} - \frac{\beta_0\hat{\beta}}{\hat{\beta}\hat{\sigma}_{\hat{\beta}}} = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}\left(1 - \frac{\beta_0}{\hat{\beta}}\right) = \frac{\hat{\beta}-\beta_0}{\hat{\beta}}.\hat{t}_{\hat{\beta}}, \tag{21}$$

so that by Proposition 5,

$$\hat{t}_{\hat{\beta},\beta=\beta_0} = \frac{\hat{\beta}-\beta_0}{\hat{\beta}}.\hat{t}_{\hat{\beta}} \sim t_{n-2}. \tag{22}$$

For example, suppose that in the ice cream case we want to test the null hypothesis $H_0: \beta = 1$. Then

$$\hat{t}_{\hat{\beta},\beta=1} = \frac{\hat{\beta}-1}{\hat{\beta}}.\hat{t}_{\hat{\beta}} = \frac{1.5-1}{1.5}\times 4.597 \approx 1.532, \tag{23}$$

which under the null hypothesis $H_0: \beta = 1$ is a random drawing from the *t* distribution with 6 degrees of freedom. Note that the value of this test statistic is in the acceptance regions in Figures 2 and 3.

This trick is useful if the econometric software you are using only reports the t-values but not the standard errors. If the standard errors are reported, you can compute $\hat{t}_{\hat{\beta},\beta=\beta_0}$ directly as $\hat{t}_{\hat{\beta},\beta=\beta_0} = (\hat{\beta}-\beta_0)/\hat{\sigma}_{\hat{\beta}}$. Of course, if only the standard errors are reported and not the t-values you can compute the t-value of $\hat{\beta}$ as $\hat{t}_{\hat{\beta}} = \hat{\beta}/\hat{\sigma}_{\hat{\beta}}$.

6. *The R^2*

The R^2 of a regression model compares the sum of squared residuals (*SSR*) of the model with the *SSR* of a "regression model" without regressors:

$$Y_j = \alpha + U_j, \quad j = 1,2,.....,n. \tag{24}$$

It is easy to verify that the OLS estimator $\tilde{\alpha}$ of α is just the sample mean of the Y_j's:

$$\tilde{\alpha} = \bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j.$$

Therefore, the *SSR* of "regression model" (24) is $\sum_{j=1}^{n}(Y_j - \bar{Y})^2$, which is called the Total Sum of Squares (*TSS*), is

$$TSS = \sum_{j=1}^{n}(Y_j - \bar{Y})^2. \tag{26}$$

The R^2 is now defined as:

$$R^2 \overset{def.}{=} 1 - \frac{SSR}{TSS}. \tag{27}$$

The R^2 is always between zero and one, because $SSR \leq TSS$. (*Exercise*: Why?) If $SSR = TSS$, so that $R^2 = 0$, then model (24) explains the dependent variable Y_j's equally well as model (2). In other words, the explanatory variables X_j in (2) do not matter: $\beta = 0$. The other extreme case is where $R^2 = 1$, which corresponds to $SSR = 0$. Then the dependent variable Y_j in model (2) is completely explained by X_j, without error: $Y_j \equiv \alpha + \beta X_j$. Thus, the R^2 measures how well the explanatory variables X_j are able to explain the corresponding dependent variables Y_j. For example, in the ice cream case, $SSR = 11.5$ and $TSS = 52$, hence $R^2 = 0.778846$. Loosely speaking, this means that about 78% of the variation of ice cream sales can be explained by the variation in temperature.

7. *Presenting regression results*

When you need to report regression results you should include, next to the OLS estimates of course, either the corresponding t-values or the standard errors, the sample size *n*, the standard error of the residuals (SER), and the R^2, because this information will enable the reader to judge your results. For example, our ice cream estimation results should be displayed as either

$$\underset{(-0.100)}{Sales = -0.25} + \underset{(4.597)}{1.5\,Temp.}, \quad n = 8, \; SER = 1.384437, \; R^2 = 0.778846$$

(t-*values between brackets*)

or

$$\underset{(2.49583)}{Sales = -0.25} + \underset{(0.32632)}{1.5\,Temp.}, \quad n = 8, \; SER = 1.384437, \; R^2 = 0.778846$$

(*standard errors between brackets*)

It is helpful to the reader if you would indicate whether you have displayed the t-values between brackets or the standard errors, but you only need to mention this once.

8. *Out-of-sample forecasting*

The linear regression model was introduced as a forecasting scheme. The question we now address is: How reliable is an out-of-sample forecast?

Consider the linear regression model (2), and suppose we observe X_{n+1}. Then the forecast of Y_{n+1} is $\hat{Y}_{n+1} = \hat{\alpha} + \hat{\beta}.X_{n+1}$, where the OLS estimators $\hat{\alpha}$ and $\hat{\beta}$ are computed on the basis of the observations for $j = 1,2,...,n$. The actual but unknown value of Y_{n+1} is

$$Y_{n+1} = \alpha + \beta.X_{n+1} + U_{n+1},$$

so that the forecast error is:

$$Y_{n+1} - \hat{Y}_{n+1} = U_{n+1} - (\hat{\alpha}-\alpha) - (\hat{\beta}-\beta).X_{n+1} = U_{n+1} - \sum_{j=1}^{n}\left(\frac{1}{n} + \frac{(X_{n+1}-\bar{X})(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right).U_j \qquad (28)$$

See the Appendix for the latter equality. It follows now from Lemma 3 that under Assumptions I through V, $Y_{n+1} - \hat{Y}_{n+1} \sim N[0, \sigma^2_{Y_{n+1}-\hat{Y}_{n+1}}]$, where

$$\sigma^2_{Y_{n+1}-\hat{Y}_{n+1}} = \sigma^2\left(\frac{n+1}{n} + \frac{(X_{n+1}-\bar{X})^2}{\sum_{j=1}^{n}(X_j-\bar{X})^2}\right). \qquad (29)$$

See the Appendix. Denoting,

$$\hat{\sigma}^2_{Y_{n+1}-\hat{Y}_{n+1}} = \hat{\sigma}^2\left(\frac{n+1}{n} + \frac{(X_{n+1}-\bar{X})^2}{\sum_{j=1}^{n}(X_j-\bar{X})^2}\right), \qquad (30)$$

it follows now similar to Proposition 6 that

Proposition 8. *Under assumptions I - V,* $(Y_{n+1} - \hat{Y}_{n+1})/\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}} \sim t_{n-2}$.

This result can be used to construct a 95% confidence interval, say, of Y_{n+1}. Look up in the table of the *t* distribution the critical value t_* of the two-sided t-test with $n-2$ degrees of freedom. Then it follows from Proposition 7 that

$$\begin{aligned} 0.95 &= P[-t_* \leq (Y_{n+1} - \hat{Y}_{n+1})/\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}} \leq t_*] \\ &= P[-t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}} \leq Y_{n+1} - \hat{Y}_{n+1} \leq t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}] \qquad (31) \\ &= P[\hat{Y}_{n+1} - t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}} \leq Y_{n+1} \leq \hat{Y}_{n+1} + t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}] \end{aligned}$$

Thus, the 95% confidence interval of Y_{n+1} is $[\hat{Y}_{n+1} - t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}, \hat{Y}_{n+1} + t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}]$.

Observe from (30) that $\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}$ increases with $(X_{n+1} - \bar{X})^2$, and so does the width of the confidence interval. Thus, the father X_{n+1} is away from $\bar{X}$, the more unreliable the forecast $\hat{Y}_{n+1}$ of Y_{n+1} becomes. Also observe from (30) that $\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}} \geq \hat{\sigma}$, and that $\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}$ gets close to $\hat{\sigma}$ if *n* is large because $\lim_{n\to\infty}\sum_{j=1}^{n}(X_j - \bar{X})^2 = \infty$.

9. *Relaxing the non-random regressor assumption*

As said before, the assumption that the regressors X_j are non-random is too strong an assumption in economics. Therefore, we now assume that the X_j 's are random variables. This requires the following modifications of the Assumptions I-V:

Assumption I*: *The pairs* (X_j, Y_j), $j = 1,2,3,....,n$, *are independent and identically distributed***.**

Assumption II*: *The conditional expectations* $E[U_j|X_j]$ *are equal to zero*: $E[U_j|X_j] \equiv 0$.

Assumption III*: *The conditional expectations* $E[U_j^2|X_j]$ *do not depend on the* X_j*'s and are finite, constant and equal*: $E[U_j^2|X_j] \equiv \sigma^2 < \infty$. (*This is called the* ***homoscedasticity*** *assumption*.)

Assumption IV*: *Conditional on X_j, U_j is $N(0,\sigma^2)$ distributed.*

The Assumptions I* and II* imply that for $j = 1,...,n$,

$$E[U_j|X_1,X_2,\ldots,X_n] \equiv 0, \tag{32}$$

and similarly the Assumptions I* and III* imply that for $j = 1,...,n$,

$$E[U_j^2|X_1,X_2,\ldots,X_n] \equiv \sigma^2. \tag{33}$$

Because (loosely speaking) conditioning on $X_1,X_2,\ldots,X_n$ is effectively the same as treating them as given constants, most of the previous propositions carry over:

Proposition 9. *Under Assumptions I*-IV*, Propositions 1 and 4 through 7 carry over, and the results in Propositions 2 and 3 now hold conditional on $X_1,X_2,\ldots,X_n$.*

However, without Assumption *IV** we need an additional condition in Proposition 6 in order to use the central limit theorem, namely:

Proposition 10. *If the sample size n is large then under the assumptions I* - III* and the additional condition $E[X_j^2] < \infty$ the approximate normality results in Proposition 7 carry over.*

Moreover, without Assumption *IV** the Propositions 6 and 8 are no longer true. As to Proposition 6, this not a big deal, as in large samples we can still use Proposition 7, but without Assumption *IV** we can no longer derive confidence intervals for the forecasts, as these confidence intervals are based on Proposition 8. It is therefore important to test the normality assumption.

10. *Testing the normality assumption*

For a normal random variable U with zero expectation and variance σ^2 it can be shown that

$$\text{Kurtosis} \overset{def.}{=} E[U^4]/\sigma^4 - 3 = 0,$$
$$\text{Skewness} \overset{def.}{=} E[U^3] = 0 \qquad (34)$$

Therefore, the normality condition can be tested by testing whether the kurtosis and the skewness of the model errors are zero, using the residuals. This is the idea behind the Jarque-Bera[3] and Kiefer-Salmon[4] tests. Under the null hypothesis (34) the test statistic involved has a χ^2_2 distribution

11. *Heteroscedasticity*[5]

We say that the errors U_j of regression model (2) are heteroskedastic if assumption III* does not hold:

$$E[U_j^2|X_j] = \psi(X_j) \textit{ for some function } \psi(.). \qquad (35)$$

Heteroscedasticity often occurs in practice. It is actually the rule rather than the exception. The main consequence of heteroscedasticity is that the conditional variance formulas in Propositions 2 and 3 do no longer hold, although the unbiasedness result in Proposition 1 is not affected by heteroscedasticity. Therefore, the Propositions 4-8 are no longer valid as well. In particular, the conditional variance of $\hat{\beta}$ [see (60)] under heteroscedasticity takes the form

$$\text{var}(\hat{\beta}|X_1,...,X_n) = E[(\hat{\beta}-\beta)^2|X_1,...,X_n] = \frac{\sum_{j=1}^{n}(X_j-\bar{X})^2\psi(X_j)}{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^2}. \qquad (36)$$

A cure for the heteroscedasticity problem is to replace the standard error of $\hat{\beta}$ by

---

[3] Jarque, C.M.and A.K. Bera, (1980), "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals". *Economics Letters* 6, 255−−259.

[4] Kiefer, N. and M. Salmon (1983), "Testing Normality in Econometric Models", *Economic Letters* 11, 123-127.

[5] Also spelled as "Heteroskedasticity."

$$\tilde{\sigma}_{\hat{\beta}} = \sqrt{\left(\frac{n}{n-2}\right)\frac{\sum_{j=1}^{n}(X_j-\bar{X})^2\hat{U}_j^2}{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^2}}. \tag{37}$$

This is known as the Heteroscedasticity Consistent (H.C.) standard error. The H.C. t-value then becomes $\tilde{t}_{\hat{\beta}} = \hat{\beta}/\tilde{\sigma}_{\hat{\beta}}$. Under the null hypothesis β = 0 this t-value is no longer *t* distributed, but the standard normal approximation remains valid if the sample size *n* is large.

A popular test for heteroscedasticity is the Breusch-Pagan[6] test. Given that

$$E[U_j^2|X_j] = g(\gamma_0 + \gamma_1 X_j) \textit{ for some unknown function } g(.). \tag{38}$$

the Breusch-Pagan test tests the null hypothesis

$$H_0\text{: } \gamma_1 = 0 \Leftrightarrow E[U_j^2|X_j] = g(\gamma_0) = \sigma^2, \textit{ say} \tag{39}$$

against the alternative hypothesis

$$H_0\text{: } \gamma_1 \neq 0 \Leftrightarrow E[U_j^2|X_j] = g(\gamma_0+\gamma_1 X_j) = \psi(X_j), \textit{ say.} \tag{40}$$

Under the null hypothesis (39) of homoskedasticity the test statistic of the Breusch-Pagan test has a χ_1^2 distribution[7], and the test is conducted right-sided.

12. *How close are OLS estimators*?

The ice cream data in Table 1 is not based on any actual observations on sales and temperature; I have picked the numbers for X_j and Y_j quite arbitrarily. Therefore, there is no way to find out how close the OLS estimates $\hat{\alpha} = -0.25$, $\hat{\beta} = 1.5$ are to the unknown parameters α and β. Actually, we do not know either whether the linear regression model (2) and its assumptions are applicable to this artificial data.

In order to show how well OLS estimators approximate the corresponding parameters I

---

[6] Breusch, T. and A. Pagan (1979), "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica* 47, 1287-1294.

[7] In the multiple regression case the degrees of freedom is equal to the number of parameters minus 1 for the intercept.

have generated random samples[8] $(Y_1,X_1),\dots,(Y_n,X_n)$ for three sample sizes: $n = 10$, $n = 100$ and $n = 1000$, as follows. The explanatory variables X_j have been drawn independently from the χ^2_1 distribution, the regression errors U_j have been drawn independently from the N(0,1) distribution, and the Y_j's have been generated by

$$Y_j = 1 + X_j + U_j,\ j = 1,2,\dots,n. \tag{41}$$

Thus, in this case the parameters α and β in model (2) are α = 1 and β = 1, and the standard error of U_j is σ = 1. Moreover, note that the Assumptions I*-IV* hold for model (41).

The true R^2 can be defined by

$$R_0^2 = 1 - \frac{E[SSR]}{E[TSS]} = 1 - \frac{(n-2)\sigma^2}{\sum_{j=1}^n E[(Y_j-\bar{Y})^2]}.$$

In the case (41), $\sigma^2 = 1$, $\mu_Y = E(Y_j) = 1 + E(X_j) = 2$,

$$\sum_{j=1}^n E[(Y_j-\bar{Y})^2] = E\left[\sum_{j=1}^n\left((Y_j-\mu_Y) - (\bar{Y}-\mu_Y)\right)^2\right] = E\left[\sum_{j=1}^n (Y_j-\mu_Y)^2 - n(\bar{Y}-\mu_Y)^2\right] = (n-1)\mathrm{var}(Y_j)$$

and

$$\mathrm{var}(Y_j) = E[(X_j - 1 + U_j)^2] = E[(X_j - 1)^2] + E[U_j^2] = E[(X_j - 1)^2] + 1 = 3,$$

because X_j is χ^2_1 distributed and therefore has the same distribution as U_j^2, and it can be shown that for standard normal random variables U_j, $E[(U_j^2-1)^2] = 2$. Thus, the true R^2 in this case is

$$R_0^2 = 1 - \frac{n-2}{3(n-1)} = \frac{2n-1}{3n-3} \approx \begin{cases} 0.7037 & for\ n = 10 \\ 0.6700 & for\ n = 100 \\ 0.6670 & for\ n = 1000 \end{cases}$$

The estimation results involved are given in Table 2:

---

[8] Via the *EasyReg International* menus File → Choose an input file → Create artificial data. Rather than generating one random sample of size $n = 1000$ and then using subsamples of sizes $n = 10$ and $n = 100$, these samples have been generates separately for $n = 10$, $n = 100$ and $n = 1000$.

Table 2: *Artificial regression estimation results*

	$\hat{\beta}$	$\hat{\alpha}$	SER (= $\hat{\sigma}$)	R^2	n
estimate:	1.11748	0.55912	0.919045	0.8842	10
(*t−value*):	(7.817)	(1.675)			
estimate:	1.03309	0.96028	0.992502	0.8284	100
(*t−value*):	(21.753)	(8.237)			
estimate:	1.02360	0.98518	0.983608	0.6899	1000
(*t−value*):	(47.124)	(26.037)			

Even for a sample size of n = 10 the OLS estimator $\hat{\beta}$ is already pretty close to its true value 1, and the same applies to $\hat{\sigma}$, but $\hat{\alpha}$ is too far away from the true value $\alpha = 1$. However, for n = 100 the OLS estimators $\hat{\beta}$ and $\hat{\alpha}$ deviate only about ±4% from their true values $\alpha = \beta = 1$, and $\hat{\sigma}$ deviates about -1% from its true value 1. In the case n = 1000 these deviations reduce to about ±2%. The R^2's are too high, and only for n = 1000 is the R^2 reasonably close to its true value. However, the R^2 is only a descriptive statistic; it does not play a role in hypotheses testing, so that the unreliability of the R^2 in small samples is harmless.

Notice the quite dramatic increase of the t-values. Recall that these t-values are the test statistics of the null hypotheses that the corresponding parameters are zero. Because the true parameters are equal to 1, what you see in Table 2 is the increase of the power of the t-test with the sample size.

APPENDIX

Proof of (1):

The first-order conditions for a minimum of $Q(\hat{\alpha},\hat{\beta}) = \sum_{j=1}^{n}(Y_j - \hat{\alpha} - \hat{\beta}X_j)^2$ are:

$$
\begin{aligned}
dQ(\hat{\alpha},\hat{\beta})/d\hat{\alpha} = 0 &\Leftrightarrow \sum_{j=1}^{n} 2(Y_j - \hat{\alpha} - \hat{\beta}X_j)(-1) = 0 \\
&\Leftrightarrow \sum_{j=1}^{n} (Y_j - \hat{\alpha} - \hat{\beta}X_j) = 0 \\
&\Leftrightarrow \sum_{j=1}^{n} Y_j - \sum_{j=1}^{n} \hat{\alpha} - \sum_{j=1}^{n} (\hat{\beta}X_j) = 0 \\
&\Leftrightarrow \sum_{j=1}^{n} Y_j = n\hat{\alpha} + \hat{\beta}\sum_{j=1}^{n} X_j = 0 \\
&\Leftrightarrow \bar{Y} = \hat{\alpha} + \hat{\beta}.\bar{X},
\end{aligned}
\tag{42}
$$

and

$$
\begin{aligned}
dQ(\hat{\alpha},\hat{\beta})/d\hat{\beta} = 0 &\Leftrightarrow \sum_{j=1}^{n} 2(Y_j - \hat{\alpha} - \hat{\beta}X_j)(-X_j) = 0 \\
&\Leftrightarrow \sum_{j=1}^{n} (Y_jX_j - \hat{\alpha}X_j - \hat{\beta}X_j^2) = 0 \\
&\Leftrightarrow \sum_{j=1}^{n} X_jY_j - \hat{\alpha}\sum_{j=1}^{n} X_j - \hat{\beta}\sum_{j=1}^{n} X_j^2 = 0 \\
&\Leftrightarrow \sum_{j=1}^{n} X_jY_j = \hat{\alpha}\sum_{j=1}^{n} X_j + \hat{\beta}\sum_{j=1}^{n} X_j^2 \\
&\Leftrightarrow \frac{1}{n}\sum_{j=1}^{n} X_jY_j = \hat{\alpha}\bar{X} + \hat{\beta}\frac{1}{n}\sum_{j=1}^{n} X_j^2
\end{aligned}
\tag{43}
$$

where $\bar{X} = (1/n)\sum_{j=1}^{n}X_j$ and $\bar{Y} = (1/n)\sum_{j=1}^{n}Y_j$ are the sample means of the X_j's and Y_j's, respectively. The last equations in (42) and (43) are called the *normal equations*:

$$\bar{Y} = \hat{\alpha} + \hat{\beta}.\bar{X}, \tag{44}$$

$$\frac{1}{n}\sum_{j=1}^{n} X_jY_j = \hat{\alpha}.\bar{X} + \hat{\beta}\frac{1}{n}\sum_{j=1}^{n} X_j^2. \tag{45}$$

To solve these normal equations, substitute $\hat{\alpha} = \bar{Y} - \hat{\beta}.\bar{X}$ in (45). Then we get

$$\frac{1}{n}\sum_{j=1}^{n} X_j Y_j = (\bar{Y} - \hat{\beta}\bar{X})\,\frac{1}{n}\sum_{j=1}^{n} X_j + \hat{\beta}\frac{1}{n}\sum_{j=1}^{n} X_j^2$$

$$= \bar{Y}.\bar{X} - \hat{\beta}\bar{X}^2 + \hat{\beta}\frac{1}{n}\sum_{j=1}^{n} X_j^2$$

$$= \bar{X}.\bar{Y} + \hat{\beta}\left(\frac{1}{n}\sum_{j=1}^{n} X_j^2 - \bar{X}^2\right)$$

hence

$$\frac{1}{n}\sum_{j=1}^{n} X_j Y_j - \bar{X}.\bar{Y} = \hat{\beta}\left(\frac{1}{n}\sum_{j=1}^{n} X_j^2 - \bar{X}^2\right). \tag{46}$$

Equation (46) can also be written as

$$\frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})(Y_j - \bar{Y}) = \hat{\beta}\left(\frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})^2\right), \tag{47}$$

because

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})(Y_j - \bar{Y}) &= \frac{1}{n}\sum_{j=1}^{n}\left(X_j Y_j - \bar{X}.Y_j - X_j.\bar{Y} + \bar{X}.\bar{Y}\right)\\ &= \frac{1}{n}\sum_{j=1}^{n} X_j Y_j - \bar{X}.\frac{1}{n}\sum_{j=1}^{n} Y_j - \bar{Y}.\frac{1}{n}\sum_{j=1}^{n} X_j + \bar{X}.\bar{Y}\\ &= \frac{1}{n}\sum_{j=1}^{n} X_j Y_j - \bar{X}.\bar{Y}\end{aligned} \tag{48}$$

and similarly

$$\frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})^2 = \frac{1}{n}\sum_{j=1}^{n} X_j^2 - \bar{X}^2. \tag{49}$$

Moreover,

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})(Y_j - \bar{Y}) &= \frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})Y_j - \frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})\bar{Y}\\ &= \frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})Y_j - (\bar{X} - \bar{X})\bar{Y}\\ &= \frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})Y_j\end{aligned} \tag{50}$$

The result (1) now follows from (44) and (46) through (50).

Proof of Proposition 1.

Recall from (1) that

$$\hat{\beta} = \frac{\sum_{j=1}^{n}(X_j - \bar{X})Y_j}{\sum_{j=1}^{n}(X_j - \bar{X})^2}. \tag{51}$$

Substitute model (2) in (51). Then

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_{j=1}^{n}(X_j - \bar{X})(\alpha + \beta X_j + U_j)}{\sum_{j=1}^{n}(X_j - \bar{X})^2} \\
&= \frac{\alpha\sum_{j=1}^{n}(X_j - \bar{X}) + \beta\sum_{j=1}^{n}(X_j - \bar{X})X_j + \sum_{j=1}^{n}(X_j - \bar{X})U_j}{\sum_{j=1}^{n}(X_j - \bar{X})^2} \\
&= \beta . \frac{\sum_{j=1}^{n}(X_j - \bar{X})X_j}{\sum_{j=1}^{n}(X_j - \bar{X})^2} + \frac{\sum_{j=1}^{n}(X_j - \bar{X})U_j}{\sum_{j=1}^{n}(X_j - \bar{X})^2} \\
&= \beta + \frac{\sum_{j=1}^{n}(X_j - \bar{X})U_j}{\sum_{j=1}^{n}(X_j - \bar{X})^2},
\end{aligned} \tag{52}$$

where the last step follows from the fact that similar to (50),

$$\sum_{j=1}^{n}(X_j - \bar{X})^2 = \sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X}) = \sum_{j=1}^{n}(X_j - \bar{X})X_j. \tag{53}$$

Now take the mathematical expectation at both sides of (52). Then,

$$E[\hat{\beta}] = \beta + E\left(\frac{\sum_{j=1}^{n}(X_j - \bar{X})U_j}{\sum_{j=1}^{n}(X_j - \bar{X})^2}\right) = \beta + \frac{\sum_{j=1}^{n}(X_j - \bar{X})E(U_j)}{\sum_{j=1}^{n}(X_j - \bar{X})^2} = \beta, \tag{54}$$

because taking the mathematical expectation of a constant (β) does not effect that constant, and taking the mathematical expectation of a linear function of random variables is equal to taking the linear function of the mathematical expectation of these random variables. The last conclusion in (54) follows from assumption II, and the second step in (54) can be taken because we have assumed that the X_j's are non-random (assumption IV).

Next consider $\hat{\alpha}$. We have already established that $\hat{\alpha} = \bar{Y} - \hat{\beta}.\bar{X}$. Substituting the right-hand side of (52) for $\hat{\beta}$ in this equation yields

$$\hat{\alpha} = \bar{Y} - \left(\beta + \frac{\sum_{j=1}^{n}(X_j - \bar{X})U_j}{\sum_{j=1}^{n}(X_j - \bar{X})^2}\right).\bar{X} = \bar{Y} - \beta.\bar{X} - \frac{\sum_{j=1}^{n}\bar{X}(X_j - \bar{X})U_j}{\sum_{j=1}^{n}(X_j - \bar{X})^2}. \quad (55)$$

Substituting

$$\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j = \frac{1}{n}\sum_{j=1}^{n}(\alpha + \beta X_j + U_j) = \alpha + \beta.\bar{X} + \frac{1}{n}\sum_{j=1}^{n} U_j$$

in (55) yields

$$\hat{\alpha} = \alpha + \frac{1}{n}\sum_{j=1}^{n} U_j - \frac{\sum_{j=1}^{n}\bar{X}(X_j - \bar{X})U_j}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \alpha + \sum_{j=1}^{n}\left(\frac{1}{n} - \frac{\bar{X}(X_j - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right).U_j. \quad (56)$$

Similar as for $\hat{\beta}$ we therefore have:

$$E[\hat{\alpha}] = \alpha + \sum_{j=1}^{n}\left(\frac{1}{n} - \frac{\bar{X}(X_j - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)E[U_j] = \alpha. \quad (57)$$

This completes the proof of Proposition 1.

Proof of Lemma 1:

We have

$$\begin{aligned} E\left[\left(\sum_{j=1}^{n} v_j U_j\right)\left(\sum_{j=1}^{n} w_j U_j\right)\right] &= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} v_i w_j U_i U_j\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} v_i w_j E(U_i U_j) \\ &= \sum_{j=1}^{n} v_j w_j \sigma^2, \end{aligned} \quad (58)$$

where the last equality in (58) follows from

$$\begin{aligned} E(U_i U_j) &= E(U_i)E(U_j) = 0 \ \ if\ i \neq j, \\ &= E(U_j^2) = \sigma^2 \ \ if\ i = j. \end{aligned} \quad (59)$$

Proof of Proposition 2:

It follows from formula (52) and Lemma 2 that

$$\begin{aligned}
\mathrm{var}(\hat{\beta}) &= E[(\hat{\beta}-\beta)^2] \\
&= E\left[\left(\sum_{j=1}^{n}\left(\frac{X_j-\bar{X}}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)U_j\right)^2\right] = \sigma^2\sum_{j=1}^{n}\left(\frac{X_j-\bar{X}}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)^2 \\
&= \sigma^2\frac{\sum_{j=1}^{n}(X_j-\bar{X})^2}{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^2} = \sigma^2\frac{\sum_{j=1}^{n}(X_j-\bar{X})^2}{\left(\sum_{j=1}^{n}(X_j-\bar{X})^2\right)^2} = \frac{\sigma^2}{\sum_{j=1}^{n}(X_j-\bar{X})^2}.
\end{aligned} \tag{60}$$

Similarly, it follows from formula (56) and Lemma 2 that

$$\begin{aligned}
\mathrm{var}(\hat{\alpha}) &= E[(\hat{\alpha}-\alpha)^2] \\
&= E\left[\left(\sum_{j=1}^{n}\left(\frac{1}{n} - \frac{\bar{X}(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)U_j\right)^2\right] = \sigma^2\sum_{j=1}^{n}\left(\frac{1}{n} - \frac{\bar{X}(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)^2 \\
&= \sigma^2\sum_{j=1}^{n}\left(\frac{1}{n^2} - \frac{2\frac{1}{n}\bar{X}(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} + \frac{\bar{X}^2(X_j-\bar{X})^2}{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^2}\right) \\
&= \sigma^2\left(\frac{1}{n} - \frac{2\bar{X}(1/n)\sum_{j=1}^{n}(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} + \frac{\bar{X}^2\sum_{j=1}^{n}(X_j-\bar{X})^2}{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^2}\right) \\
&= \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{j=1}^{n}(X_j-\bar{X})^2}\right) \\
&= \sigma^2\left(\frac{(1/n)\sum_{j=1}^{n}(X_j-\bar{X})^2 + \bar{X}^2}{\sum_{j=1}^{n}(X_j-\bar{X})^2}\right) = \frac{\sigma^2\sum_{j=1}^{n}X_j^2}{n\sum_{j=1}^{n}(X_j-\bar{X})^2},
\end{aligned} \tag{61}$$

where the last equality follows from the fact that $(1/n)\sum_{j=1}^{n}(X_j-\bar{X})^2 = (1/n)\sum_{j=1}^{n}X_j^2 - \bar{X}^2$.

Finally, it follows from Lemma 1 and the formulas (52) and (56) that

$$\text{ov}(\hat\alpha,\hat\beta) = E[(\hat\alpha-\alpha)(\hat\beta-\beta)] = E\left[\left(\sum_{j=1}^{n}\left(\frac{1}{n} - \frac{\bar X(X_j-\bar X)}{\sum_{i=1}^{n}(X_i-\bar X)^2}\right)U_j\right)\left(\sum_{j=1}^{n}\left(\frac{X_j-\bar X}{\sum_{i=1}^{n}(X_i-\bar X)^2}\right)U_j\right)\right]$$

$$= \sigma^2\sum_{j=1}^{n}\left(\frac{1}{n} - \frac{\bar X(X_j-\bar X)}{\sum_{i=1}^{n}(X_i-\bar X)^2}\right)\left(\frac{(X_j-\bar X)}{\sum_{i=1}^{n}(X_i-\bar X)^2}\right) \tag{62}$$

which can be rewritten as

$$\text{cov}(\hat\alpha,\hat\beta) = \sigma^2\left(\frac{(1/n)\sum_{j=1}^{n}(X_j-\bar X) - \bar X\sum_{j=1}^{n}(X_j-\bar X)^2}{\left(\sum_{i=1}^{n}(X_i-\bar X)^2\right)^2}\right) = \frac{-\sigma^2.\bar X}{\sum_{j=1}^{n}(X_j-\bar X)^2}. \tag{63}$$

Proof of Proposition 5.

Observe first from (44) and (9) that

$$\frac{1}{n}\sum_{j=1}^{n}\hat U_j = \bar Y - \hat\alpha - \hat\beta.\bar X = 0 \tag{64}$$

so that we can write

$$\hat U_j = \hat U_j - \frac{1}{n}\sum_{i=1}^{n}\hat U_i = (Y_j - \bar Y) - \hat\beta.(X_j-\bar X). \tag{65}$$

Next, observe from (2) that $Y_j - \bar Y = U_j - \bar U + \beta.(X_j-\bar X)$, where $\bar U = (1/n)\sum_{j=1}^{n}U_j$. Substituting the former equation in (65) yields

$$\hat U_j = (U_j - \bar U) - (\hat\beta-\beta)(X_j-\bar X), \tag{66}$$

hence

$$\sum_{j=1}^{n}\hat U_j^2 = \sum_{j=1}^{n}\left((U_j-\bar U) - (\hat\beta-\beta)(X_j-\bar X)\right)^2$$

$$= \sum_{j=1}^{n}(U_j-\bar U)^2 - 2(\hat\beta-\beta)\sum_{j=1}^{n}(X_j-\bar X)(U_j-\bar U) + (\hat\beta-\beta)^2\sum_{j=1}^{n}(X_j-\bar X)^2 \tag{67}$$

$$= \sum_{j=1}^{n}(U_j-\bar U)^2 - 2(\hat\beta-\beta)\sum_{j=1}^{n}(X_j-\bar X)U_j + (\hat\beta-\beta)^2\sum_{j=1}^{n}(X_j-\bar X)^2,$$

where the last equality follows from the fact that $\sum_{j=1}^{n}(X_j-\bar{X})\bar{U} = 0$. It follows from (52), (67) and the equality $\sum_{j=1}^{n}(U_j-\bar{U})^2 = \sum_{j=1}^{n}U_j^2 - n\bar{U}^2$ that

$$\begin{aligned}\sum_{j=1}^{n}\hat{U}_j^2 &= \sum_{j=1}^{n}(U_j-\bar{U})^2 - (\hat{\beta}-\beta)^2\sum_{j=1}^{n}(X_j-\bar{X})^2 = \sum_{j=1}^{n}U_j^2 - n\bar{U}^2 - (\hat{\beta}-\beta)^2\sum_{j=1}^{n}(X_j-\bar{X})^2.\\ &= \sum_{j=1}^{n}U_j^2 - \frac{1}{n}\left(\sum_{i=1}^{n}U_i\right)^2 - (\hat{\beta}-\beta)^2\sum_{j=1}^{n}(X_j-\bar{X})^2.\end{aligned} \tag{68}$$

Taking expectations and using Lemma 2 and Proposition 2 it follows now from (68) that

$$\begin{aligned}E[\textstyle\sum_{j=1}^{n}\hat{U}_j^2] &= \textstyle\sum_{j=1}^{n}E[U_j^2] - \dfrac{1}{n}E\left[\left(\sum_{i=1}^{n}U_i\right)^2\right] - \left(E(\hat{\beta}-\beta)^2\right)\sum_{j=1}^{n}(X_j-\bar{X})^2\\ &= n\sigma^2 - \sigma^2 - \sigma^2 = (n-2)\sigma^2.\end{aligned} \tag{69}$$

Proof of (13):

$$\begin{aligned}SSR &= \sum_{j=1}^{n}\hat{U}_j^2 = \sum_{j=1}^{n}(Y_j - \hat{\alpha} - \hat{\beta}.X_j)^2 = \sum_{j=1}^{n}(Y_j - (\bar{Y}-\hat{\beta}.\bar{X}) - \hat{\beta}.X_j)^2\\ &= \sum_{j=1}^{n}\left((Y_j - \bar{Y}) - \hat{\beta}.(X_j-\bar{X})\right)^2\\ &= \sum_{j=1}^{n}(Y_j - \bar{Y})^2 - 2\hat{\beta}\sum_{j=1}^{n}(Y_j - \bar{Y})(X_j-\bar{X}) + \hat{\beta}^2\sum_{j=1}^{n}(X_j-\bar{X})^2\\ &= \sum_{j=1}^{n}(Y_j - \bar{Y})^2 - \hat{\beta}^2\sum_{j=1}^{n}(X_j-\bar{X})^2.\end{aligned} \tag{70}$$

Proof of (28):

It follows from (3) that

$$\begin{aligned}Y_{n+1}-\hat{Y}_{n+1} &= U_{n+1} - (\hat{\alpha}-\alpha) - (\hat{\beta}-\beta).X_{n+1}\\ &= U_{n+1} - \sum_{j=1}^{n}\left(\frac{1}{n} - \frac{\bar{X}(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right).U_j - \sum_{j=1}^{n}\left(\frac{X_{n+1}(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)U_j\\ &= U_{n+1} - \sum_{j=1}^{n}\left(\frac{1}{n} + \frac{(X_{n+1}-\bar{X})(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right).U_j.\end{aligned} \tag{71}$$

Proof of (29):

It follows from (28) and Lemma 3 that

$$\sigma^2_{Y_{n+1}-\hat{Y}_{n+1}} = \sigma^2 + \sum_{j=1}^{n}\left(\frac{1}{n} + \frac{(X_{n+1}-\bar{X})(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)^2.\sigma^2$$

$$= \sigma^2\left(1 + \frac{1}{n} + \frac{2}{n}.\frac{(X_{n+1}-\bar{X})\sum_{j=1}^{n}(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} + \frac{(X_{n+1}-\bar{X})^2\sum_{j=1}^{n}(X_j-\bar{X})^2}{(\sum_{i=1}^{n}(X_i-\bar{X})^2)^2}\right) \tag{72}$$

$$= \sigma^2\left(\frac{n+1}{n} + \frac{(X_{n+1}-\bar{X})^2}{\sum_{j=1}^{n}(X_j-\bar{X})^2}\right).$$