

2 ECONOMETRIC METHODS

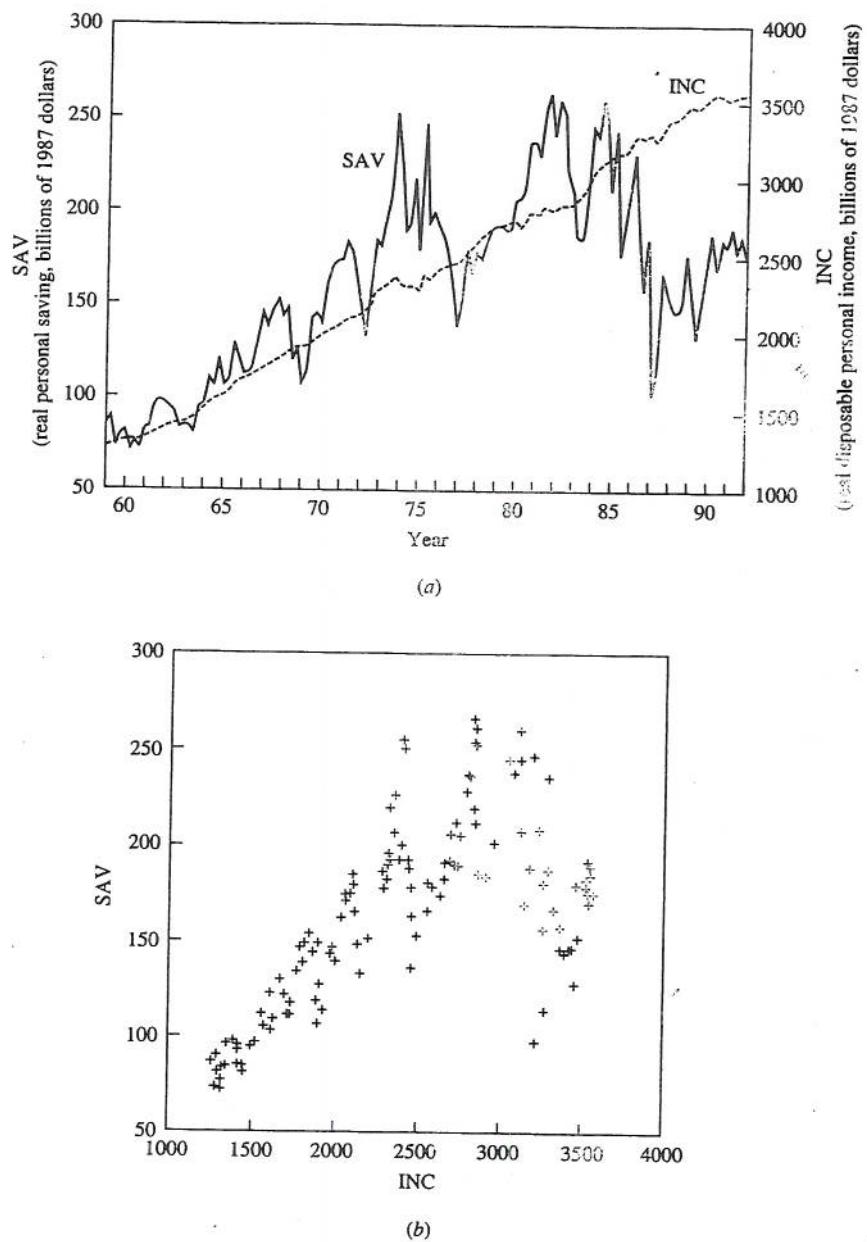


FIGURE 1.1
Saving and income.

4 ECONOMETRIC METHODS

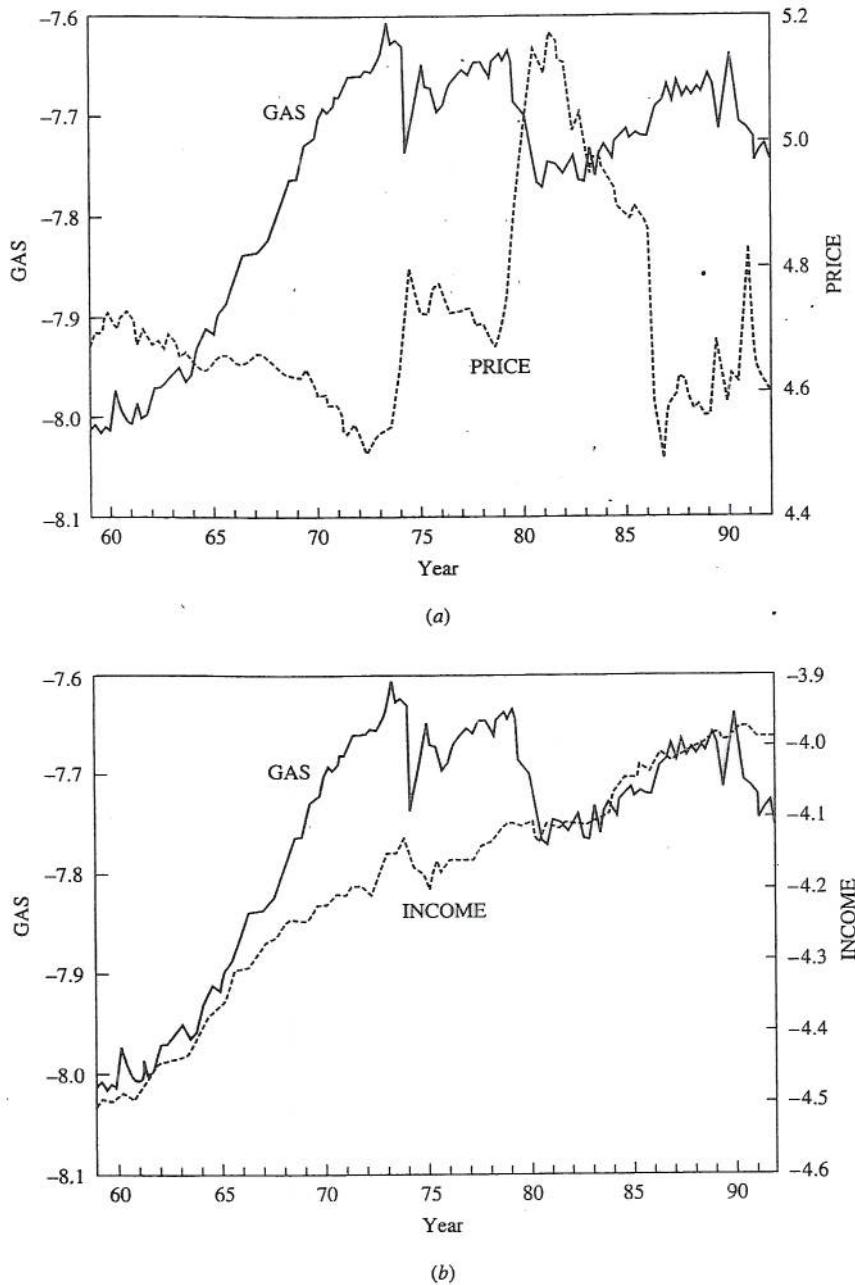
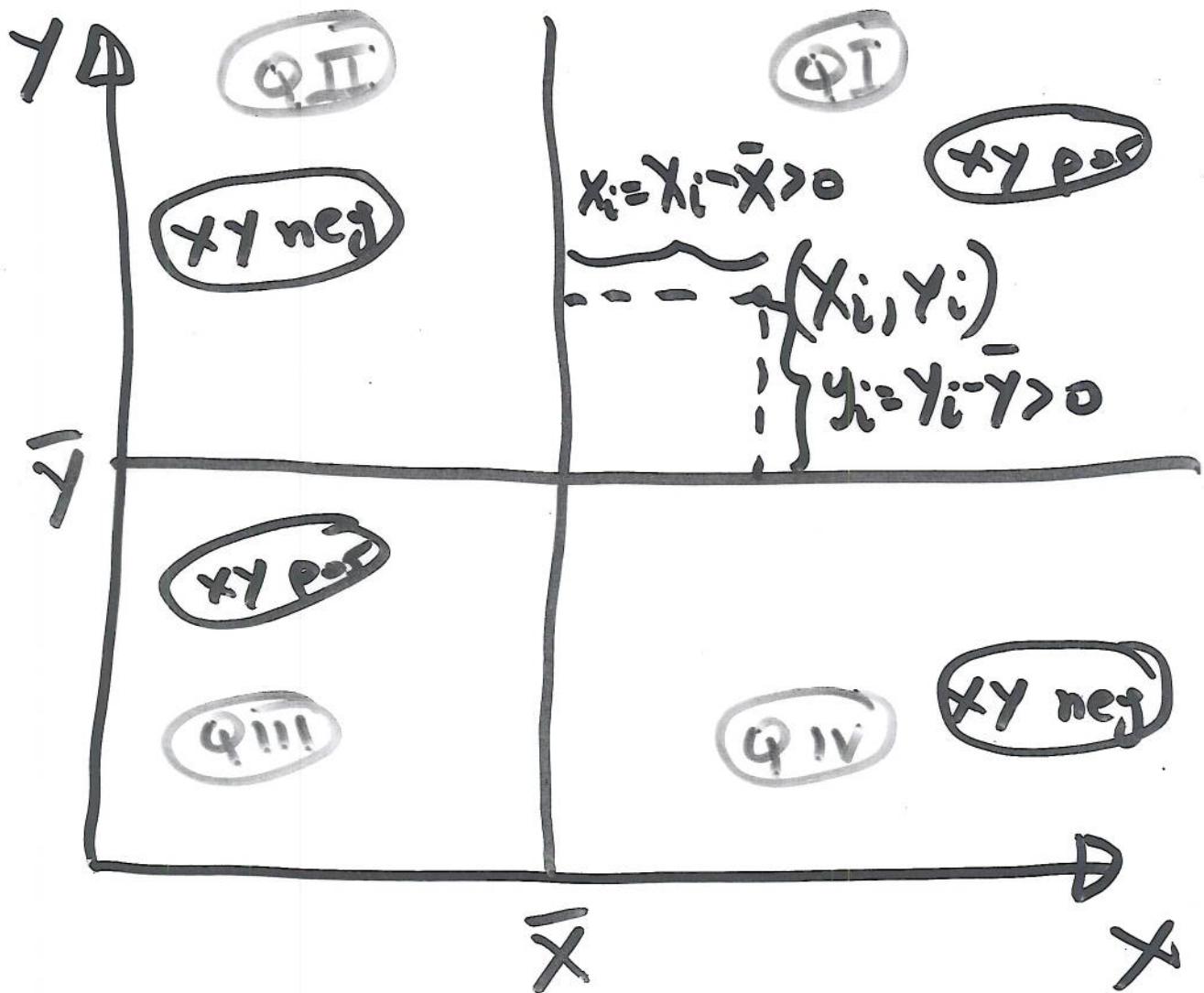


FIGURE 1.2

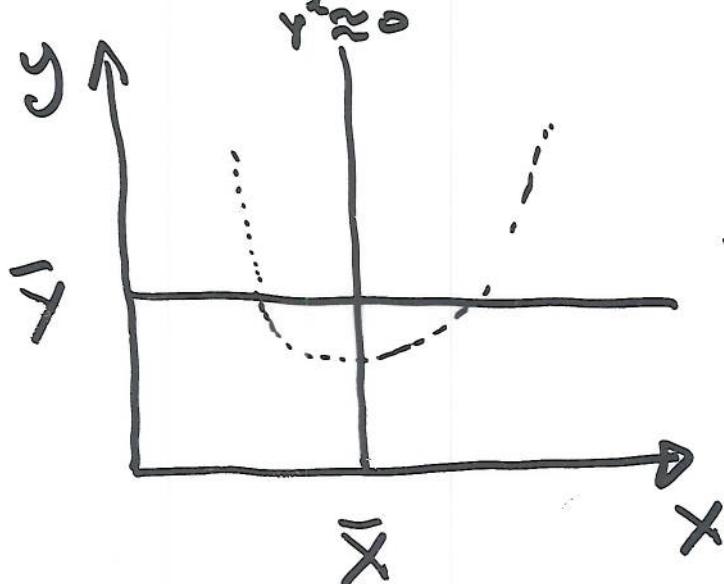
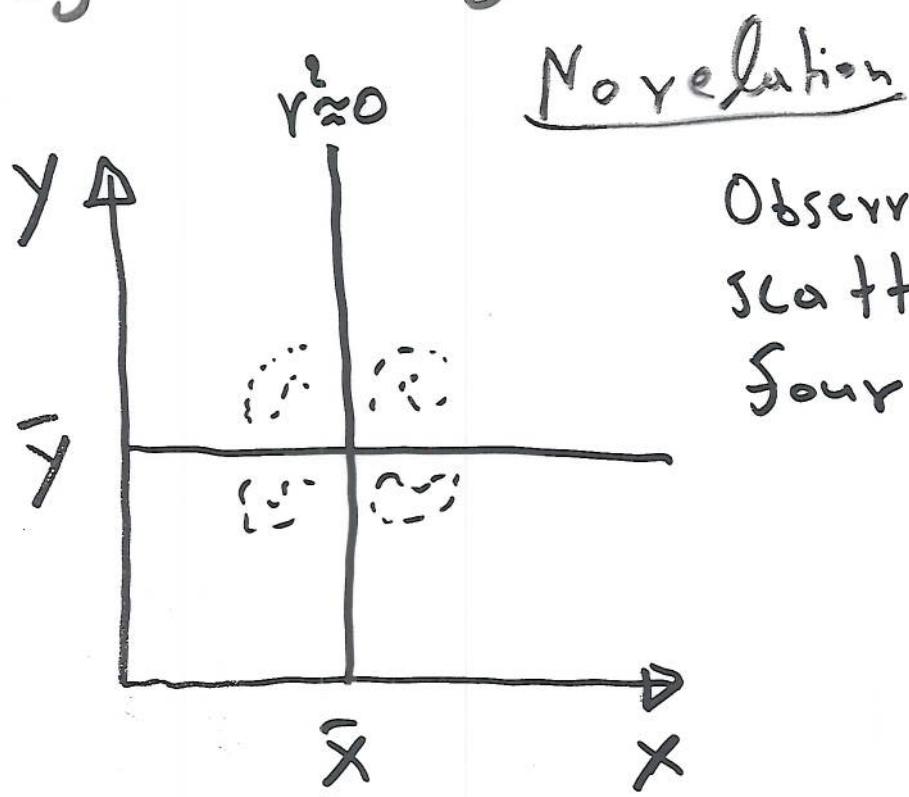
Time series plots of natural log of gasoline consumption in 1987 dollars per capita.
(a) Gasoline consumption vs. natural log of price in 1987 cents per gallon. (b)
Gasoline consumption vs. natural log of income in 1987 dollars per capita.

SCATTER DIAGRAM



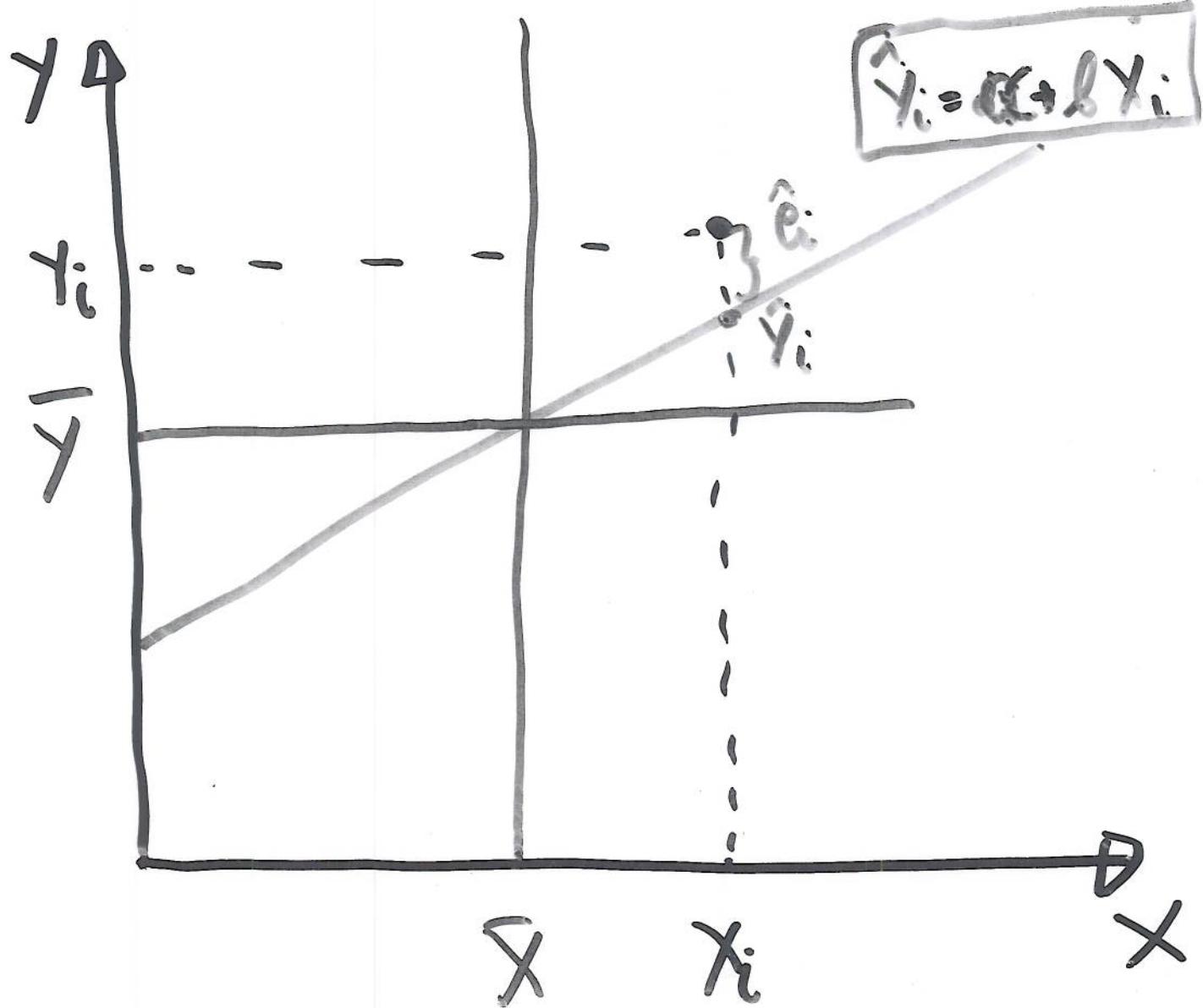
(5)

(Correlation coefficients must be interpreted with care. Many coefficients that are both numerically large and also adjudged statistical by tests may contain no



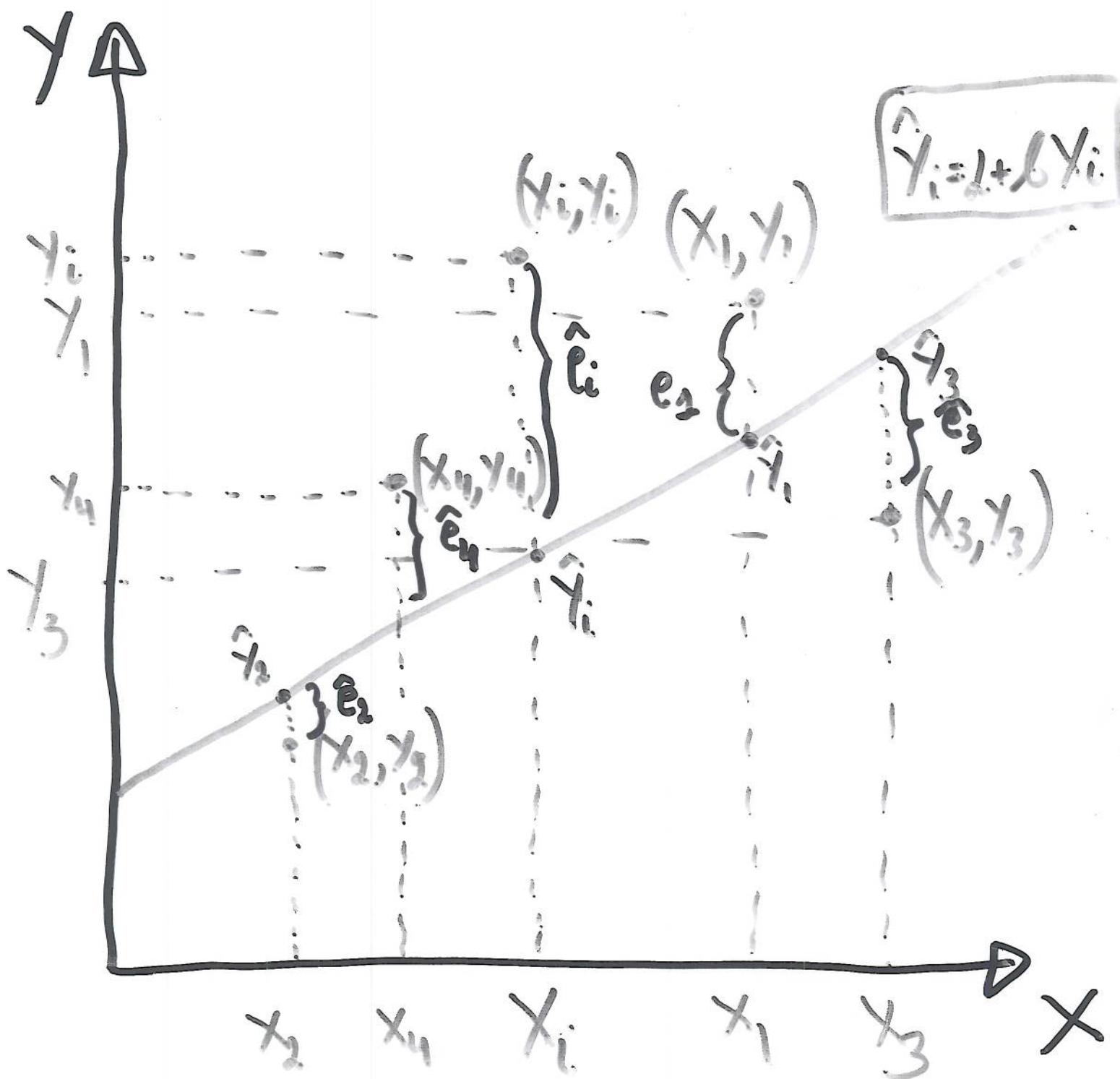
LEAST-SQUARE ESTIMATORS

Residual: $\hat{e}_i = Y_i - \hat{Y}_i = Y_i - a - bX_i \quad i=1, 2, \dots, N$



Early years of the 19th century:
least squares → a dominant and powerful
estimating principle

* Select a, b to minimize the sum of squared errors: $\sum_{i=1}^n e_i^2$



$$e_i = y_i - \hat{y}_i = y_i - (a + b x_i)$$

CHAPTER 1: Relationships between Two Variables

s in the Y variable
of squares from the regression of Y

the regression of Y on X
give
 $\frac{\text{ESS}}{\text{TSS}}$ (1.34)

f the Y variation attributable to the
s an alternative demonstration that
case the sample points all lie on a

application of these formulae. Sub-

b
20b

X
able 1.7. The regression coefficients

$$= 1.75$$

$$.75(4) = 1$$

as

$$(70) = 122.5$$

traction as

$$1 - 122.5 = 1.5$$

ned by the linear regression is

$$= 0.9879$$

quares. Some authors use SSR to indicate the
o indicate the sum of squares due to error (one

TABLE 1.6

X	Y	XY	X^2	\hat{Y}	e	xe
2	4	8	4	4.50	-0.50	-1
3	7	21	9	6.25	0.75	2.25
1	3	3	1	2.75	0.25	0.25
5	9	45	25	9.75	-0.75	-3.75
9	17	153	81	16.75	0.25	2.25
Sums	20	40	230	120	40	0

TABLE 1.7

x	y	xy	x^2	y^2	\hat{y}	e	xe
-2	-4	8	4	16	-3.50	-0.50	1.00
-1	-1	1	1	1	-1.75	0.75	-0.75
-3	-5	15	9	25	-5.25	0.25	-0.75
1	1	1	1	1	1.75	-0.75	-0.75
5	9	45	25	81	8.75	0.25	1.25
Sums	0	0	70	40	124	0	0

DEVIATION
from THE
MEAN

1.5 INFERENCE IN THE TWO-VARIABLE, LEAST-SQUARES MODEL

The least-squares (LS) estimators of α and β have been defined in Eqs. (1.28) to (1.30). There are now two important questions:

What are the properties of these estimators?

How may these estimators be used to make inferences about α and β ?

1.5.1 Properties of LS Estimators

The answers to both questions depend on the **sampling distribution** of the LS estimators. A sampling distribution describes the behavior of the estimator(s) in repeated applications of the estimating formulae. A given sample yields a specific numerical estimate. Another sample from the same population will yield another numerical estimate. A sampling distribution describes the results that will be obtained for the estimator(s) over the potentially infinite set of samples that may be drawn from the population.

The parameters of interest are α , β , and σ^2 of the conditional distribution, $Y|X$. In that conditional distribution the only source of variation from one hypothetical sample to another is variation in the stochastic disturbance (u), which in association with the given X values will determine the Y values and hence the sample estimates of α , β , and σ^2 . Analyzing Y conditional on X thus treats the X_1, X_2, \dots , as fixed in repeated sampling. This treatment rests on the implicit assumption

contents of the probability statement may be rearranged to give a 95 percent confidence interval for σ^2 as

$$\frac{(n-2)s^2}{\chi^2_{0.975}} \quad \text{to} \quad \frac{(n-2)s^2}{\chi^2_{0.025}}$$

1.5.4 Numerical Example (Continued from Section 1.4.5)

From the data in Tables 1.6 and 1.7 we have already calculated

$$\begin{array}{lll} n = 5 & a = 1 & b = 1.75 \\ \text{TSS} = 124 & \text{ESS} = 122.5 & \text{RSS} = 1.5 \quad r^2 = 0.9879 \end{array}$$

We now obtain

$$\begin{aligned} s^2 &= \text{RSS}/(n-2) = 1.5/3 = 0.5 \\ \text{var}(b) &= s^2 / \sum x^2 = 0.5/40 = 0.0125 \\ \text{var}(a) &= 0.5 \left(\frac{1}{5} + \frac{16}{40} \right) = 0.3 \end{aligned}$$

The estimated standard errors of the regression coefficients are thus

$$\text{s.e.}(a) = \sqrt{0.3} = 0.5477 \quad \text{s.e.}(b) = \sqrt{0.0125} = 0.1118$$

A preselected critical value from the t distribution with 3 degrees of freedom is $t_{0.025} = 3.182$. Thus, a 95 percent confidence interval for a is

$$1 \pm 3.182(0.5477)$$

that is,

$$-0.74 \quad \text{to} \quad 2.74$$

and a 95 percent confidence interval for b is

$$1.75 \pm 3.182(0.1118)$$

that is,

$$1.39 \quad \text{to} \quad 2.11$$

The intercept is not significantly different from zero since

$$\frac{a}{\text{s.e.}(a)} = \frac{1}{0.5477} = 1.826 < 3.182$$

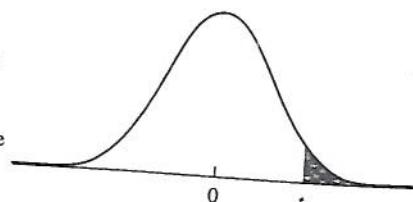
whereas the slope is strongly significant since

$$\frac{b}{\text{s.e.}(b)} = \frac{1.75}{0.1118} = 15.653 > 3.182$$

As indicated earlier, once confidence intervals have been computed, actually computing the significance tests is unnecessary, since a confidence interval that includes zero is equivalent to accepting the hypothesis that the true value of the parameter

TABLE D.2
Student's *t* distribution

The first column lists the number of degrees of freedom (*v*). The headings of the other columns give probabilities (*P*) for *t* to exceed the entry value. Use symmetry for negative *t* values.



<i>v</i>	<i>P</i>	.10	.05	.025	.01	.005
1		3.078	6.314	12.706	31.821	63.657
2		1.886	2.920	4.303	6.965	9.925
3		1.638	2.353	3.182	4.541	5.841
4		1.533	2.132	2.776	3.747	4.604
5		1.476	2.015	2.571	3.365	4.032
6		1.440	1.943	2.447	3.143	3.707
7		1.415	1.895	2.365	2.998	3.499
8		1.397	1.860	2.306	2.896	3.355
9		1.383	1.833	2.262	2.821	3.250
10		1.372	1.812	2.228	2.764	3.169
11		1.363	1.796	2.201	2.718	3.106
12		1.356	1.782	2.179	2.681	3.055
13		1.350	1.771	2.160	2.650	3.012
14		1.345	1.761	2.145	2.624	2.977
15		1.341	1.753	2.131	2.602	2.947
16		1.337	1.746	2.120	2.583	2.921
17		1.333	1.740	2.110	2.567	2.898
18		1.330	1.734	2.101	2.552	2.878
19		1.328	1.729	2.093	2.539	2.861
20		1.325	1.725	2.086	2.528	2.845
21		1.323	1.721	2.080	2.518	2.831
22		1.321	1.717	2.074	2.508	2.819
23		1.319	1.714	2.069	2.500	2.807
24		1.318	1.711	2.064	2.492	2.797
25		1.316	1.708	2.060	2.485	2.787
26		1.315	1.706	2.056	2.479	2.779
27		1.314	1.703	2.052	2.473	2.771
28		1.313	1.701	2.048	2.467	2.763
29		1.311	1.699	2.045	2.462	2.756
30		1.310	1.697	2.042	2.457	2.750
40		1.303	1.684	2.021	2.423	2.704
60		1.296	1.671	2.000	2.390	2.660
120		1.289	1.658	1.980	2.358	2.617
∞		1.282	1.645	1.960	2.326	2.576

Reprinted from P. B. Hoel, *Introduction to Mathematical Statistics*, 4th edition, Wiley, 1971, by permission of the publishers.