12-1-1958

Economic Theory of Teams. Chapter 3

Jacob Marschak

Roy Radner

COWLES FOUNDATION DISCUSSION PAPER NO. 59c

Economic Theory of Teams*

Chapter 3

J. Marschak and R. Radner

March 2, 1959

* Research undertaken by the Cowles Commission for Research in Economics under Contract Nonr-358(01), NR 047-006 with the Office of Naval Research

ECONOMIC THEORY OF TEAMS

Chapter 3

EXAMPLES

1. Example A: Buying faultless market information. A firm suffers a loss if it either underestimates or overestimates the demand for its product. Assume this loss to be about proportional to the absolute value of the error: i.e., loss $= k \cdot |x-a|$, $k > 0$, where x is the true demand and a is the amount (called supply) that the firm brings to the market and that is equal to its estimate of demand. The firm knows the probability distribution of demand: x can be small ($= 1$), medium ($= 2$) or large ($= 3$), with probabilities .1, .3, and .6, respectively. The firm is faced with the following alternatives:

1. to determine the supply on the basis of its own knowledge of the probabilities of demand;

2. to pay a market research agency which we shall suppose faultless and which provides, for different fees, the following kinds of information: it can tell whether the demand will be

   2' : small or not small
   2'': large or not large
   2''': medium or not medium

3. to pay the market research agency for information on whether the demand will be small, medium, or large.

Problem: What are the minimum expected losses under each of the five alternatives (1; 2', 2'', 2'''; 3), not counting the research fees? How much should the firm be willing to pay, at most, for each of the four research services, if it tries to maximize its expected profit, or in other words, to minimize its expected losses?

Since the main purpose of the example is to illustrate certain abstract concepts, let us perform a translation. The five cases are identified with five different information structures which can be numbered in the same way. Thus $\eta = 1, 2', 2'', 2''', \text{ or } 3$. To each of the five values of η corresponds a different set Y of subsets of X (a different partition of X), which we represent, by enclosing each relevant subset of X (i.e., each element of Y) into brackets:

$\eta = 1$; Y = ([1, 2, 3]); (one subset of X)

$\eta = 2'$; Y = ([1], [2,3]); (two subsets of X)

$\eta = 2''$; Y = ([1,2], [3]); (two subsets of X)

$\eta = 2'''$; Y = ([1,3], [2]); (two subsets of X)

$\eta = 3$; Y = ([1], [2], [3]); (three subsets of X)

We can compute and compare the minimal expected losses, $\Omega(\eta)$,*

* In our list of concepts, section 2..., the minimal expected loss was denoted by $\hat{\Omega}(\eta)$; but we can here omit the ^ symbol, for brevity.

for each of the five information structures. First let us conveniently tabulate our payoff function (or, rather, loss function: we shall minimize, not maximize, and thus save cumbersome minus signs), $\omega(x,a)$, and the probabilities $\varphi(x)$ of the states of nature. The quantities x and a are expressed in appropriate units so that $k = 1$; $|x-a|$ = loss.

Supply (a) \ Demand (x)	1	2	3
1	0	1	2
2	1	0	1
3	2	1	0
Probabilities	.1	.3	.6

In case 1, the expected losses $E\omega(x, a)$ for each of the three actions are:

$$E\omega(x,1) = (0)(.1) + (1)(.3) + (2)(.6) = 1.5$$

$$E\omega(x,2) = (1)(.1) + (0)(.3) + (1)(.6) = .7$$

$$E\omega(x,3) = (2)(.1) + (1)(.3) + (0)(.6) = .5$$

The best decision is to constantly maintain $a = 3$, resulting in an average loss of .5. Thus the minimum expected loss $\Omega(\eta)$ is equal to .5 when $\eta = 1$. We write $\Omega(1) = .5$.

For the case 2', first compute the minimal expected losses conditional upon each of the two possible communications obtained: $x = 1$ and $x \neq 1$, respectively. Then compute the weighted average of the two conditional expectations (see Section 2.5 above). When $x = 1$, the optimal $a = 1$, minimal loss = 0. When $x \neq 1$, the optimal a is the one that gives the smallest of the following expected losses (use columns 2 and 3 of the above table):

$$(1)\cdot(3/9) + (2)\cdot(6/9) = 5/3 \text{ (when } a = 1)$$

$$(0)\cdot(3/9) + (1)\cdot(6/9) = 2/3 \text{ (when } a = 2)$$

$$(1)\cdot(3/9) + (0)\cdot(6/9) = 1/3 \text{ (when } a = 3)$$

Hence best $a = 3$, yielding the minimal conditional expected loss 1/3. Since $x = 1$ occurs with probability .1, and $x \neq 1$ with probability .9, we have

$$\omega(2') = (0)(.1) + (1/3)(.9) = .3$$

By similar operations we find:

$$\omega(2'') = (0)(.6) + \min(3/4, 1/4, 5/4)\cdot(.4) = .1$$

$$\omega(2''') = (0)(.3) + \min(12/7, 1, 2/7)\cdot(.7) = .2$$

Finally if the research agency identifies the demand precisely, then always the optimal $a = x$, $\omega(x, a) = 0$; hence, $\Omega(3) = 0$.

Summarizing our results:

information structure η	minimum expected loss $\Omega(\eta)$	value of information structure $\Omega(1) - \Omega(\eta)$
1	.5	0
2'	.3	.2
2''	.1	.4
2'''	.2	.3
3	0	.5

The ranking of the figures in the second column agrees with the statements made in Section 2.8. Information structure $\eta = 3$ is finer than any of the structures 2', 2'', 2''', and, accordingly, is at least as profitable as any one of these; information structure 1 is less fine than either 2' or 2'' or 2''' and, accordingly, is not more profitable than any one of these. No comparison of fineness can be made among 2', 2'', 2'''; the ranking of their expected loss will vary with the parameters of the problem.

The lowest profit being, under all conditions, associated with information structure $\eta = 1$, the values of information structures, in column 3, are as defined in Section 2.7. The column gives the upper bounds on the research fees that the firm should be willing to pay for each kind of service.

2. Example B: The speculator. Suppose a speculator cannot buy or sell more than one unit of a commodity. Let x be the difference between tomorrow's and today's price and suppose x is distributed uniformly over

the interval $[-1, +1]$; or, speaking somewhat loosely: x can take, with equal probability, any value between -1 and $+1$. Our problem is again to compare the maximum expected payoffs possible under several alternative information structures. This time, each will be characterized by a different number of sub-intervals of equal length into which the whole interval $[-1, +1]$ is partitioned. Each of these information structures can be unambiguously labelled as $n = 1, 2, \ldots\ldots$ We have to evaluate $\Omega(n)$ for various integers n.

Denote by a the amount bought (if $a \geq 0$) or sold (if $a \leq 0$). Then $-1 \leq a \leq 1$, and the payoff $u = ax$.

Let $n = 1$; i.e., the speculator is not informed about the amount or direction of the price change. Since $Eax = aEx$ and he knows that $Ex = 0$ (in fact this is the only relevant a priori information, in this case), all decisions a based on this information yield the same expected payoff -- viz., zero -- and are therefore equally good. Hence $\Omega(1) = 0$.

Let $n = 2$: the speculator knows whether the price will rise or fall; he will buy or sell, accordingly. That is: if $x > 0$, the best decision $a = 1$; if $x < 0$, $a = -1$ (and if $x = 0$ any a will be equally good). The payoff yielded by this decision rule is $u = |x|$; the conditional expected payoff, in the case of rising as well as of falling prices, is equal to $1/2$ (= mid-value between 0 and 1) and since each of the two cases is equally probable, the expected payoff $\Omega(2) = 1/2$.

Let $n = 4$; that is, the speculator is informed, not only whether the price will rise or fall, but also whether it will change by more or by less than 1/2 unit. Clearly this additional information will not change the best decision rule: to buy (sell) when the price is going to rise (fall); at each x, the payoff will remain $u = |x|$; and the expected payoff $\Omega(4) = \Omega(2) = 1/2$. This will remain so no matter in how many subintervals the positive and the negative part of the interval [-1, +1] is partitioned. Hence $\Omega(2k) = 1/2$ for any positive integer k. It is also clear that if x is always exactly known to the decision maker, this will not change the decision rule just given, nor add to the expected payoff of 1/2. We can say that $\Omega(\infty) = 1/2$.

But now let $n = 3$. Although the problem remains simple, we shall introduce here the information variable y explicitly, to illustrate our system of concepts more fully. y will now have 3 values, which we can denote thus:

$$y = \begin{matrix} y_- \\ y_o \\ y_+ \end{matrix} \quad \text{if } x \text{ is in the interval} \quad \begin{matrix} [-1, -1/3] \\ [-1/3, 1/3] \\ [1/3, 1] . \end{matrix}$$

Clearly if $y = y_+$ or y_- the best actions are $\hat{\alpha}(y) = 1$ or -1, respectively, and $u = |x|$; the conditional expected payoffs are in each of these two cases equal to 2/3 (mid-value between 1/3 and 1). But if $y = y_o$, any action yields the same conditional expected payoff 0 (analogous to the case $n = 1$ above); thus

$$E[\omega(x, \hat{\alpha}(Y))|y] = \begin{Bmatrix} 2/3 \\ 0 \\ 2/3 \end{Bmatrix} \quad \text{if } y = \begin{cases} y_+ \\ y_o \\ y_- \end{cases}$$

$\Omega(3)$ is the weighted average of these three quantities (with equal weights 1/3); hence $\Omega(3) = 4/9$.

The fact that $\Omega(1) < \Omega(2) = \Omega(4) = \Omega(2k)$ (with k any integer) $= \Omega(\infty)$ agrees with the "if" part of the statement of Section 2.8: that, with any payoff function, making the information structure finer (in the sense defined) never decreases but possibly increases the expected payoff. Moreover, our payoff function happens to be such that $\Omega(3) < \Omega(2)$. This is in agreement with the "only if" part of the statement. For, in the sense defined, neither of the structures "2" and "3" is finer than the other (although the latter carries a larger "amount of information" in Shannon's sense). And our payoff conditions make it more important to distinguish between, say, 1/5 and -1/5 (which, in information structure "3" result in the same information y_o but in the case "2" result in different informations) than to distinguish between 1/5 and 2/5 (which is possible under "3" but impossible under "2"). In fact, dividing the interval [-1, 1] into any odd number, however large, of equal sub-intervals, will be always less valuable to our decision maker than dividing into just two sub-intervals, the positive and the non-positive: this is due to the fact that, when the number of sub-intervals is odd, the knowledge that x has fallen into the middle sub-interval is of no value to the speculator who needs to know whether x is positive or negative.

3. Example C: Production with constant returns. Suppose one unit of input produces one unit of output, and the plant's capacity is limited to one unit. Denote the price of output by $m_1 + x_1$, the price of input by $-(m_2 + x_2)$ where each x_i is random with zero mean. We shall assume the expected profit $(m_1 + m_2)$ to be equal to zero, and each random price

deviation x_i $(i = 1,2)$ to have only two values, x_i = "high" = s_i (> 0) or x_i = "low" = $-s_i$, with equal probabilities; assume the probabilities of each of the four possible pairs to be:

(1)

$x_1 =$ \ $x_2 =$	$-s_2$	$+s_2$
$-s_1$	$(1+r)/4$	$(1-r)/4$
$+s_1$	$(1-r)/4$	$(1+r)/4$

where $-1 \leq r \leq 1$. Hence the conditional probabilities are (for $i \neq j$)

$$\Pr(x_i \text{ high} | x_j \text{ high}) = \Pr(x_i \text{ low} | x_j \text{ low}) = (1 + r)/2 \tag{2}$$
$$\Pr(x_i \text{ low} | x_j \text{ high}) = \Pr(x_i \text{ high} | x_j \text{ low}) = (1 - r)/2$$

It is easily seen that the above distribution implies, for each i, $Ex_i = 0$, $Ex_i^2 = s_i^2$ and $Ex_1x_2 = rs_1s_2$. Hence s_i is the standard deviation of x_i, and r is the correlation coefficient. Without loss of generality, we shall assume $s_1 \geq s_2$. Our problem will be to inquire how the distribution parameters s_1, s_2, r influence the expected payoffs of the 4 information structures denoted as follows:

η = [00]; neither x_1 nor x_2 are known;

[10]; x_1 but not x_2 is known;

[01]; x_2 but not x_1 is known;

[11]; x_1 and x_2 are both known.

We shall denote $\Omega(\eta)$ by U_η.

Denote output (= input) by a; then $0 \leq a \leq 1$ and the profit $u = a(x_1 + x_2)$.

Let η = [00]. $Eu = 0$ for any a; therefore $\Omega(00) = 0$.

Let η = [11]. Then the sum $x_1 + x_2$ is known exactly and a good decision rule (comparable to that of the case "$\eta = \infty$" of Example B) is:

$a = 1$ if $x_1 + x_2 > 0$, $a = 0$ otherwise. But since $s_1 \geq s_2$, $x_1 + x_2 > 0$ implies $x_1 = s_1$. Hence it suffices to observe x_1; that is, $U_{11} = U_{10}$, the decision rule being: $a = 1$ if x_1 is high, $a = 0$ otherwise. Using the conditional probabilities (2) already computed above, we obtain the following conditional expected profits that this rule will yield:

$$E\left\{1 \cdot (x_1 + x_2) | x_1 = s_1\right\} = [(s_1 - s_2)(1-r) + (s_1 + s_2)(1 + r)]/2$$

$$= s_1 + rs_2$$

$$E\left\{0 \cdot (x_1 + x_2) | x_1 = -s_1\right\} = 0.$$

Since $Pr(x_1 = s_1) = 1/2$, $\Omega(10) = \Omega(11) = (s_1 + rs_2)/2^a$, positive quantity.

Finally, let $\eta = [01]$. Using again our table of conditional probabilities (2) we obtain

$$E(u|x_2 \text{ high}) = [a \cdot (s_1 + s_2)(1+r) + a(-s_1 + s_2)(1-r)]/2 = a(rs_1 + s_2)$$

$$E(u|x \text{ low}) = [a (-s_1 - s_2)(1+r) + a(s_1 - s_2)(1-r)]/2 = -a(rs_1 + s_2)$$

The expression in the second line is never positive and is therefore maximized at $a = 0$: any good decision rule will require to stop production when x_2 is low. But the expression in the first line has an ambiguous sign: it can be positive or non-positive according as the correlation coefficient r exceeds or does not exceed the negative fraction $- s_2/s_1$. Hence the good decision rule will depend on the parameters. If $r \leq \frac{-s_2}{s_1}$, no production is ever worthwhile [because, roughly, a low (high) input price is likely to be associated with low (high) output price]; in this $U_{01}\ U_{00} = 0$. If, however, r exceeds the critical non-positive number $-s_2/s_1$, (which can be as low as -1 or as high as 0 depending on the relative variability of the two prices), then the good decision rule will prescribe production at full capacity ($a = 1$) provided x_2 is high; and no production if x_2 is low. This rule will result in

$$U_{01} = (rs_1 + s_2)/2 ,$$

a quantity that is always larger than U_{00} but never larger than U_{10}.

To summarize: $U_{00} = 0$; $U_{01} = \max\ (0, rs_1 + s_2)/2$; $U_{10} = (s_1 + rs_2)/2 = U_{11}$.

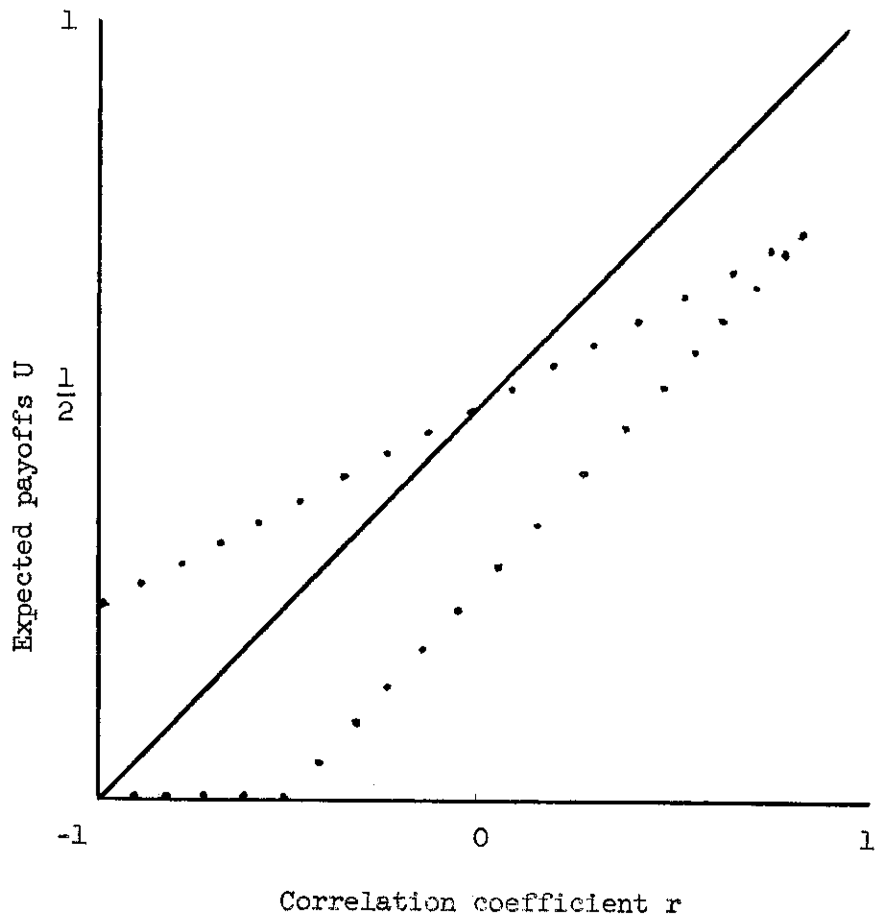
The payoffs U_{11} (which is equal to U_{10}) and U_{01} are plotted against the correlation coefficient r on the following chart, where $s_1 = 1$ and $s_2 = 1/2$ or 1. Since $U_{00} = 0$ the graphs give also the values of the information structures. There is never an advantage in observing the less volatile of the two variables over observing the more volatile one. The advantage of observing the more volatile one (x_1) approaches zero as the variances tend to become equal. The advantage of observing the more volatile variable also increases as the positive correlation approaches 1; for, roughly speaking, x_1 can then be predicted from x_2 with decreasing error. However, prediction might also be made if correlations were strongly negative. One might expect the difference $U_{10} - U_{01}$ to be symmetrical with respect to the correlation coefficient, with the maximum at $r = 0$. This is, however, not the case: the difference between payoffs of information structures depends on the payoff function and not on the parameters of the probability distribution.

To conclude, note that the information amounts (Shannon measures) -- say, $i(\eta)$ -- of the considered structures obey the following relation:

$$i(11) > i(10) = i(01) > i(00);$$

on the other hand, the payoffs were shown to obey the relations:

$$\Omega(11) = \Omega(10) > \Omega(01) > \Omega(00);$$



Graph of U_{11} ($= U_{10}$) and U_{01}

When $s_2 = 1/2$, U_{11} is given by the upper dotted line, U_{01} by the lower dotted line. When $s_2 = 1$, $U_{11} = U_{10} = U_{01}$ are all given by the solid line.

finally, the relations of fineness were;

[11] is finer than both [10] and [01];

both [10] and [01] are finer than [00];

[10] and [01] not comparable in fineness.

This confirms again the statements of Section 2.8.

4. Example D. A case of decreasing returns: output a quadratic function of a single input.

This model is adapted from traditional economists who, with a good instinct, exploited the simple mathematical properties of the case when the marginal productivity of an input smoothly diminishes as the amount of input increases ("law of diminishing returns"). We make the case more specific by using a quadratic approximation. Denote by b the single input (or, more generally, of a bundle of inputs which can be applied only in constant proportions); and assume output $\psi(b)$ to be quadratic in b, with a negative second derivative (this is implied by diminishing marginal productivity). Then it is possible to choose input units so as to make the quadratic term in $\psi(b)$ equal to -1:

$$\psi(b) = -b^2 + gb + h,$$

say. Assume the output to have a constant price. By appropriate choice of output units it can be made = 1. Denote by m the mean of input unit price, and by $m + x$ the current input unit price. Then the profit is quadratic in b:

$$u = \psi(b) \cdot 1 - (m + x)b;$$

$$(1) \quad u = -b^2 + (g - m)\, b + h - bx.$$

This can be further simplified without loss of generality, by measuring input from an appropriate origin as follows. Denote by $a = b-b^*$ the deviation of input from a certain constant, $b^* = (g - m)/2$ (the economic meaning of this constant will become apparent presently). Then $b = a + b^*$ and the profit (1) can be rewritten as

$$(2) \quad u = \omega(x, a) = -a^2 - ax + u^* - b^*x$$

where u^* is another constant. Clearly if $x = 0$ then the profit u has a unique maximum at $a = 0$; i.e., when the input b is equal to b^*. Thus the constant b^*, the new origin we have chosen, is the input that is optimal when the input price is at its mean level; and the constant u^* is the maximum profit that is then obtained. The payoff function ω depends on the environment variable x (input price measured from its mean level) and the action variable a (input measured from that level which is the best one at the mean input price).

The terms $(u^* - b^*x$ in Equation (2) are of little interest since they do not depend on the decision variable $\underline{a}$. The same value of $\underline{a}$ that maximizes the profit $\underline{u}$ maximizes also the profit measured as a deviation from $(u^* - bx)$. We can therefore re-define the origin from which profits are measured, and express the profit thus measured (thus changing the meaning of the symbols u and ω in a trivial way and making maximum profit at mean price equal to zero) as

$$(3) \quad u = \omega(x,a) = -a^2 - ax .$$

Our problem is to derive the best decision functions, and measure the resulting expected profits, under various alternative information structures. We shall consider two information structures:

1) $\eta(x) = x$, i.e., the producer is kept informed of the price;

2) $\eta(x) = X$, i.e., the producer is not so informed (the set X comprises all real members).

As in Example B we may call the first information structure, "∞" and the second "1". We shall denote the maximum expected profits, $\hat{\Omega}(\infty)$ and $\hat{\Omega}(1)$ by U_∞ and U_1, respectively.

If $\eta = \infty$ the producer will, upon learning the value of x, choose the input $\hat{a}$ that maximizes u for that value of x. Setting the derivative of (3) equal to zero (thus "equating the marginal product to the price of input") we obtain as the optimal decision, $\hat{a} = -x/2$. The maximum profit is $\hat{u} = x^2/4$. Since the expectation of x is zero the maximum expected payoff

(4) $U_\infty = Ex^2/4 = s^2/4$

where s^2 is the variance of x.

On the other hand, if $\eta = 1$, so that the producer does not know x, the best action will be some constant independent of x. It is obtained by maximizing, with respect to a, the expected profit $Eu = -a^2$. The optimal input is now $\hat{a} = 0$, yielding the maximum profit $\hat{u} = 0$, a constant. Hence the maximum expected payoff $E\hat{u}$ under information structure $\eta = 1$ is

$$U_1 = 0.$$

The advantage of the information structure "∞" over "1", i.e., the advantage of being kept informed about the current price of input,

(5) $U_\infty - U_1 = s^2/4$

In terms of Section 2.7 this is the value of the information structure "∞", i.e., the value of getting exact information about x.

The result (5) seems to agree with common sense: the advantage of knowing the value of a variable should be the larger, the less "predictable" it is, or the larger its "variability." However, variance is not the only possible measure of the vague property "variability"; variance happens to be relevant in our particular case, with the payoff function a quadratic one.*

---

* One might use a different economic illustration of a quadratic payoff function, also adapted from ancient mathematical economics (Cournot). A monopolist chooses a price $\underline{a}$ of his product so as to maximize the profit $u = a\psi(a) - c$ where $\psi(a)$ is the quantity demanded at price $\underline{a}$, and $\underline{c}$ is the total cost, assumed constant. Assume the demand function ψ linear. Then (using appropriate units of measurement) $\psi(a) = -a + m + x$ where $m + x$ is the intercept of the demand curve and m the average value of the intercept; thus, x measures the random "shift" in the public's desire for the product. If x is known the best decision rule is $\hat{\alpha}(x) = (m + x)/2$; if x is not known, the best rule is $\hat{\alpha}(x) = m/2$. The value of information about x is again proportional to the variance of x.

It is also worthy of note that the example given in the text extends to the case of imperfect market. Suppose the price of input is linear function, $m + x + ka$, where x is a random "shift" with zero mean. Then equation (1) still applies, with the coefficients properly re-interpreted.

---

5. Example E. Output a quadratic function of two inputs. We shall generalize Example D to the case when two inputs have to be used, and the producer can freely vary their quantities. This will bring out the role of an important characteristic of the payoff function, the "complementarity" or "interaction" between various actions. The distinction between payoff functions with and without interaction will prove of great importance in the theory of teams, but is already present in the case of single-person decisions. Moreover, the example will throw additional light on the role of correlation between states of the world, already discussed in Example C.

Let x_i ($i = 1, 2$) denote the price of the i-th input, measured from the mean level of that price. Suppose as before that the output price is constant; set it equal to 1 by a proper choice of the units in which output is measured. Suppose the output is a quadratic function of the two inputs. It is possible to measure inputs in such units, and from such origins, as to express the profit thus:

(1) $u = \omega(x, a_1 = a_2) = -a_1^2 - a_2^2 + 2qa_1a_2 - a_1x_1 - a_2x_2 + u* +$

terms linear in x_1 and x_2 and independent of a_1 and a_2 .

This is analogous to Equation (2) of Example D. Again, if both prices are at their mean levels, $x_1 = x_2 = 0$, then the maximum profit (equal to u*) is attained at $a_1 = a_2 = 0$.* This is the economic interpretation of the

---

* For, if $x_1 = x_2 = 0$, then (1) can be rewritten thus:

$$u = u* -(a_1^2 - 2qa_1a_2 + q^2a_2^2) + q^2a_2^2 - a_2^2 = u* - (a_1 - qa_2) = a_2^2 (1 - q^2).$$

The term $-(a - qa_2)^2$ is smallest (=0) when $a_1 = qa_2$; and the term $-a_2^2 (1-q^2)$ is smallest when $a_2 = 0$, since we have assumed $-1 < q < 1$, $q^2 < 1$. Hence when $a = a = 0$, the profit attains its maximum value, u*. Thus, the condition $-1 < q < 1$ guarantees that the "profit surface" has a summit. A more general method of proof is referred to in Chapter 5.

---

term u* and of the origins chosen for input measurements. Moreover, as in Equation (3) of Example D, it is convenient to measure the profit from an origin chosen so as to get rid of the terms that are affected by the action variables a_1 and a_2 . With this new definition of u (and ω),

(2) $u = \omega (x, a_1, a_2) = -a_1^2 -a_2^2 + 2qa_1a_2 - a_1x_1 - a_2x_2$

As in Example C, the state of nature is described by two variables (x_1, x_2), giving rise to the same four information structures [00], [10], [01],[11]; but the action that was described by a single variable in Example C, will now have two dimensions, a_1 and a_2. The economic meaning of the parameter q of the payoff function follows from the fact that

$$(3) \qquad -q/2 = \frac{\partial}{\partial a_1}\left(\frac{\partial u}{\partial a_2}\right) = \frac{\partial}{\partial a_2}\left(\frac{\partial u}{\partial a_1}\right) = \frac{\partial^2}{\partial a_1 \, \partial q_2} \; .$$

Thus q is proportional to the (second) cross-derivative of the payoff with respect to a_1 and a_2 ; it measures the extent to which the marginal contribution of a_1 is affected by the level of a_2; and also the extent to which the marginal contribution a_2 is affected by the level of a_1. A word used for this measure in older economic literature was "complementarity" between two factors of production; it can be positive, negative or zero. In the work of J. R. Hicks [ ] the term complementarity has become attached to a property of the utility (or of the production) function which is mathematically somewhat more complicated. To avoid confusion we shall use the word "interaction." When (and only when) $q = 0$, the payoff can be represented as a sum of a function of a_1 and a function of a_2:

$$(4) \qquad \omega(x, a_1, a_2) = \omega_1(x, a_1) + \omega_2(x, a_2),$$

where

$$(5) \qquad \omega_i(x, a_i) = -a_i^2 - a_i x_i = u_i \; .$$

More generally, for a payoff function of any form, we shall say that there is no interaction between its several action variables if the payoff can be represented as a sum of independent separate contributions of each variable, as in (4). When the payoff function has second derivatives (as in the quadratic case) we can, in addition, measure the degree of interaction, as in (3).

We now proceed to find the optimal decision rules and the resulting expected payoffs, under each of the four formation structures considered.

If $\eta = [00]$, i.e., no information about prices is gathered and actions are "routine", the expected profit is equal to

$$Eu = - a_1^2 - a_2^2 + 2qa_1a_2 + 0 + 0 .$$

which is to be maximized with respect to a_1 and a_2. The optimal actions are constant, $\hat{a}_1 = \hat{a}_2 = 0$, and the maximum expected profit is

(6) $$U_{00} = 0.$$

Because of (6), the value $U_\eta - U_{00}$ of any information structure η, will be simply equal to the expected payoff U_η.

If $\eta = [11]$, i.e., both prices are known before the decision is made, the optimal inputs $\hat{a}_1, \hat{a}_2$ are obtained by maximizing the profit $\omega(x_1, x_2, a_1, a_2)$ separately with respect to a_1 and a_2, and equating each partial derivative to 0; this will result in a maximum profit because of the condition $-1 < q < 1$, as shown in the preceding footnote.

(7) $$\begin{aligned} -2\hat{a}_1 + 2q\hat{a}_2 &= x_1 \\ 2q\hat{a}_1 - 2\hat{a}_2 &= x_2 , \end{aligned}$$

i.e., the "marginal product of each input should equal its price." This result is, of course, well known from the economics of certainty. Solving (7) we obtain two decision functions, each linear in x_1 and x_2:

(8) $$\begin{aligned} \hat{a}_1 &= - 1/2 \cdot \frac{x_1 + qx_2}{1 - q^2} \\ \hat{a}_2 &= -1/2 \cdot \frac{qx_1 + x_2}{1 - q^2} \end{aligned}$$

Thus the optimal quantity of an input falls with its price; and it falls (rises) with the price of the other input if q is positive (negative). If $q = 0$, each optimal input depends on its own price only (in fact, we obtain again the result of Example D).

Substituting (8) into (2) we obtain, for given x_1, x_2, the maximum profit

$$\omega(x_1, x_2, \hat{a}_1, \hat{a}_2) = \frac{x_1^2 + 2qx_1x_2 + x_2^2}{4(1 - q^2)} \quad ; \tag{9}$$

The expected maximum profit is therefore

$$U_{11} = \frac{s_1^2 + 2qrs_1s_2 + s_2^2}{4(1 - q^2)} \tag{10}$$

where s_i^2 is the variance of x_i and r is the correlation coefficient; for by definition, since $Ex_1 = Ex_2 = 0$, those parameters are

$$s_i^2 = E(x_i - Ex_i)^2 = Ex^2 ,$$

$$r = E(x_1 - Ex_1)\ (x_2 - Ex_2)/s_1s_2 = Ex_1x_2/s_1s_2$$

The quantity $U_{11}(= U_{11} - U_{00})$ measures the advantage of decisions taken in full knowledge of both prices, over mere "routine" actions. Equation (10) shows that this advantage depends not only on the variances of the prices -- compare the result of Example D -- but also on their correlation, provided there is interaction. This advantage is the larger the larger the product qr of the interaction and the correlation coefficients. Hence if the correlation is positive but the interaction is negative the advantage of using information about both prices (as against using information on none) is smaller than if interaction and correlation are both positive or both negative.

To compare the value of knowing both prices with the value of knowing only one, and to compare the value of knowing only x_1 with that of knowing only x_2, we have to consider the remaining information structures: [10] and [01].

Consider the case $\eta = [10]$: x_1, but not x_2, is known when the decision is made. We have to maximize, with respect to a_1 and a_2, the conditional expected profit

(11) $$E[u|x_1] = -a_1^2 - a_2^2 + 2qa_1a_2 - a_1x_1 - a_2\, E[x_2|x_1];$$

equating to zero the partial derivatives of this expression with respect to a_1 and a_2 we obtain

(12) $$\begin{aligned} -2\hat{a}_1 + 2q\hat{a}_2 &= x_1 \\ 2q\hat{a}_1 - 2\hat{a}_2 &= E[x_1|x_2]; \end{aligned}$$

i.e., "the marginal product of each input should equal the conditional expectation of its price," a statement more general than the one we used after equation (7).

At this point, we have to make some specific assumptions about the probability distribution of (x_1, x_2) to explicitly evaluate $E[x_2|x_1]$. It will be convenient to assume a normal distribution,* because then

---

* In Chapter 5, in an example of a quadratic payoff to a 2-person <u>team</u>, a simple discrete distribution (that of Example C above) will be used. Some of the results are applicable to the 1-person problem.

$E[x_2|x_1]$ is a linear function of x_1:

(13) $$E[x_2|x_1] = x_1 \cdot rs_1/s_2;$$

(rs_1/s_2 is the "regression coefficient of x_2 on x_1"). Substituting (13) into (12) and solving, we obtain two decision functions, each linear in x_1, the one price that is known to the decision-maker:

(14)
$$\hat{a}_1 = \frac{-x_1}{2(1 - q^2)} \cdot (1 + qr \frac{s_2}{s_1})$$

$$\hat{a}_2 = \frac{-x_1}{2(1 - q^2)} \cdot (q + r \frac{s_2}{s_1})$$

We note that the decision about a_2 is simply the routine decision, $\hat{a}_2 = 0$, (i.e., the knowledge of x_1 remains unused in determining a_2), if there is neither correlation between x_1 and x_2, nor interaction between a_1 and a_2, i.e., if $q = r = 0$. This clearly makes sense. It also makes sense that, if $q = 0$ but $r \neq 0$, $\hat{a}_2$ does depend on x_1. For, although the profit can then be decomposed into independent sub-profits due to each of the two inputs separately, as in (4), and in addition, the sub-profit u_2 due to the second input depends only on the second price, as in (5), information about x_1 does help to increase u_2 because that knowledge contains some information about the correlated variable x_2.

Inserting (14) and (13) into equation (11) one finds the best conditional expected payoff, given x_1. Taking the expected value of the resulting expression one obtains after some tedious algebra

(15) $$U_{10} = \frac{s_1^2 + s_2^2 r^2 + 2qrs_1 s_2}{4(1 - q^2)}$$

For the case $\eta = [01]$, one obtains $\hat{a}_1, \hat{a}_2$ (this time as functions of x_2 only) interchanging the subscripts 1 and 2 in (14); and by a similar interchange of subscripts in (15), one gets the expected payoff

$$U_{01} = \frac{s_2^2 + s_1^2 r^2 + 2qrs_1 s_2}{4(1 - q^2)} \tag{16}$$

It is interesting to compare U_{10} and U_{01}. Subtracting:

$$U_{10} - U_{01} = \frac{(s_1^2 - s_2^2)\ (1 - r^2)}{4\ (1 - q^2)} \tag{17}$$

The formula shows that if the two prices have equal variances (making the conditions symmetrical with respect to the inputs measured in appropriate units), it is equally useful to know only x_1 or only x_2. Also, if the two prices are strongly correlated, either positively or negatively (so that the one can be estimated from the other without too large an error), it does not matter much which one of the two prices is known. On the other hand the advantage of knowing the more volatile rather than the more constant price is the larger the stronger is the (positive or negative) interaction.

What is the advantage of knowing both prices over that of knowing only one, say x_1? From equations (10) and (15),

$$U_{11} - U_{10} = \frac{s_2^2(1 - r^2)}{4(1 - q^2)} \tag{18}$$

Thus the advantage of adding information about x_2 to that about x_1 is the larger the stronger the interaction between a_1 and a_2. Moreover this advantage is proportionate to $s_2^2(1 - r^2)$, the square of the so called "standard error of estimating x_2 from x_1."

A more complete discussion becomes possible if the costs of information are known. To illustrate, let c be the cost of getting information about either x_1 or x_2; and let 2c be the cost of getting information on both. Since

$$(19) \qquad U_{11} \geq \max (U_{10}, U_{01}) \geq U_{00}$$

the best information structure is [11] when c = 0 and [00] when c is very large. One may ask whether, for some intermediate c , either the information structure [10] or [01] (depending on whether s_1 or s_2 is larger) can be optimal. This depends on whether we have "decreasing returns to information" in the special sense that

$$(20) \qquad U_{11} - U_{10} \leq U_{10} - U_{00}$$

The latter condition is satisfied [see (6), (10), (18)] if and only if

$$(21) \qquad r\,(q + r) > 0 \,.$$

For simplicity, let us assume $s_1 = s_2$. The net expected payoffs are

$$V_{00} = U_{00} = 0$$

$$V_{10} = U_{10} - c = -c + \frac{(1 + r^2 + 2\,qr)s^2}{4\,(1 - q^2)}$$

$$V_{11} = U_{11} - 2c = -2c + \frac{2(1 + qr)s^2}{4(1 - q^2)}$$

Denote by k the cost redefined in new units, $k = \frac{4c\,(1 - q^2)}{s^2}$. Then the condition for [10] to be optimal is

$$(22) \qquad 1 - r^2 \leq k \leq 1 + r^2 + 2qr$$

But if (21) is not satisfied, $1 + r^2 + 2qr < 1 - r^2$, and no value k will satisfy (22); thus, in absence of "decreasing returns" in the sense of (20), the intermediate information structure [10] cannot be optimal: either both or none of the variables x_1, x_2 have to be observed.

In a more general case, the cost of information may not double with the number of variables observed; for example it will less than double if the information involves some fixed cost. The general method remains to compute and compare the net expected payoffs of each of the information structures considered.

6. Example F: Buying inaccurate information. Modify Example D as follows: the producer learns, not the exact price of a unit of input, but only some forecast of it, which may exceed or fall short of the actual price by a random error. The state x of the world is now described by a pair of variables, $x = (x_1, x_2)$ where x_1 will now denote the true value of the price variable, measured from its mean; and x_2 will denote the forecaster's error.* The information obtained by the producer is

---

* See discussion at the end of Section 2.2.

---

(1) $$\eta(x) = y = x_1 + x_2$$

We may assume without loss of generality that the error has no bias, i.e.,

(2) $$Ex_2 = 0 \; ;$$

for, if the bias is known the decision maker can always allow for it and correct the information accordingly.

Various methods of information gathering will presumably result in greater or smaller accuracy, measured by some kind of average of the errors of estimation. To fix the ideas we shall assume the price x_1, and the error x_2 to be independent and normally distributed. This will make it possible (as we shall see) to characterize each information structure by a single parameter, the variance of error. As in the previous examples, our problem consists in comparing the expected payoffs generated by different information structures.

The profit is (as in Example D)

(3) $$u = -a^2 - ax_1 + \text{terms independent of } a ,$$

where a is measured from the level that is best at $x_1 = 0$. We shall now recast the problem in a form that is particularly simple and also has the advantage of being identical with a well-known problem in statistics. Since the terms in (3) that are independent of $\underline{a}$ do not influence the optimal value of $\underline{a}$, and since we can change the units of measurement of price* so as to replace x_1 by $2x_1$, we can redefine the payoff function as

* The units of input have been already fixed, so as to make the coefficient of the square term in (3) equal to -1; the change in the price scale implies therefore a change in units of money which, of course, we are still free to make.

(4) $$u = -a^2 -2ax_1 + x_1^2 = -(a - x_1)^2$$

Clearly $(a - x_1)^2$ has its minimum, and u has its maximum, at $a = x_1$. Our producer is thus in the same position as a statistician whose "decision" consists in naming an estimate (a) of a quantity (x_1), and who incurs a penalty (loss) proportional to the square of the difference between the true value x_1 and his estimate a. The assumption of a quadratic loss function has been widely used in statistics. In later parts of the present book, we shall exploit the analogy between economic choice and the decision of a statistician in more complicated contexts.

If no information is obtained, the best action is that constant a that maximizes the expectation

$$(5) \qquad Eu = -a^2 + 2aEx_1 - Ex_1^2 = -a^2 - Ex_1^2$$

since $Ex_1 = 0$. Clearly the best action is

$$(6) \qquad \hat{a} = 0$$

Thus the maximum expected profit obtainable without information is

$$(7) \qquad U = -s_1^2 ,$$

where s_1^2 is the variance of x_1. This is, of course, the same result as in equation (4) of Example D, account taken of the changes in the units of x_1 and the origin of u.

On the other hand: if information y (= $x_1 + x_2$) is obtained, the decision maker maximizes with respect to $\underline{a}$ the conditional expectation

$$(8) \qquad E[u|y] = -E[(a-x_1^2)|y]$$

to obtain the value $\hat{a} = \hat{\alpha}(y)$ that is best at given y; that is, he minimizes the conditional expected loss

$$-E[u|y] = a^2 - 2aE[x_1|y] + E[x_1^2|y] ;$$

the last term does not involve $\underline{a}$ and can be neglected. We minimize therefore

$$a^2 - 2aE[x_1|y] = a^2 - 2aE[(y-x_2)|y]$$

$$= a^2 - 2a\ (y - E[x_2|y]) \ .$$

The value of a that minimizes this expression is

(9) $$\hat{a} = y - E[x_2|y]$$

To make the evaluation of $E[x_2|y]$ simple we shall narrow down our assumption. Assume both x_1 and x_2 to be normally distributed and denote their variances by s_1 and s_2, both positive. Then x_2 and y are jointly normally distributed, and

(10) $$E[x_2|y] = y.\ \text{(regression coefficient of } x_2 \text{ on } y)$$

$$= y \cdot \frac{\text{covariance of } x_2 \text{ and } y}{\text{variance of } y} \ ;$$

Since $Ex_2 = Ey = 0$,

$$E[x_2|y] = y \ \frac{Ex_2\ y}{Ey^2} \ = y \cdot \frac{Ex_2(x_2 + x_1)}{E(x_2 + x_1)^2}$$

For further simplicity, assume x_1 and x_2 to be non-correlated. Then $Ex_1x_2 = 0$ and

$$E[x_2|y] = y \cdot \frac{s_2^2}{s_2^2 + s_1^2} \ ;$$

inserting this in (9) gives

(11) $$\hat{a} = y\ (1 - \frac{s_2^2}{s_2^2 + s_1^2}) \ .$$

Hence, under our assumptions, the best rule of estimating x_1 from inaccurate information y is to correct the information by "scaling down" the absolute value of y, thus bringing it closer to the mean value (zero) of x_1. The correction is the stronger, the larger the variance (s_2^2) of the error x_2 relative to the variance of x_1 itself. As s_2^2 increases $\hat{a}$ approaches zero; this corresponds to equation (6), the case of "no information."

The profit yielded by this rule at a given state of nature (x_1, x_2) is:

$$-(\hat{a} - x_1)^2 = - [(x_2 + x_1) \frac{s_1^2}{s_1^2 + s_2^2} - x_1]^2$$

$$= -(x_2 s_1^2 - x_1 s_2^2)^2/(s_1^2 + s_2^2)^2$$

The expected profit (averaged over all states of nature) is (remembering that $Ex_1x_2 = 0$)

$$U = - (s_2^2 s_1^4 + s_1^2 s_2^4)/(s_1^2 + s_2^2)^2$$

(12)
$$U = - \frac{s_1^2 s_2^2}{s_1^2 + s_2^2} = \frac{-1}{(1/s_1^2) + (1/s_2^2)}$$

Thus the expected payoff is the smaller the smaller each of the two variances. Moreover as s_2^2, the variance of the error, increases, U approaches $-s_1^2$, as in equation (7), the case of "no information." On the other hand, as s_2^2 approaches zero, i.e., information becomes more accurate, we approach

(13) $U = 0$;

The difference between equations (13) and (7) is in accordance with equation (5) of Example D.

Given the variance s_1^2 of the price, the expected payoff U as expressed in (12) is a function of s_2^2, the variance of the error: $U = \Omega(s_2)$, say. Suppose one forecaster offers information with error variance v', at cost c'; and another offers information with error variance v'' at cost c'', then the first forecaster is preferred if

$$\Omega(v') - c' > \Omega(v'') - c'' .$$