Random Sampling in Measurement of Net National Product

By *S. Zagoroff*,
Food Research Institute, Stanford University, USA.

The two basic concepts of national income statistics on which the following analysis is based are the *gross value of output* (i. e. selling value of output including indirect taxes) and the *net product* (i. e. net value of output = value added) of a productive enterprise.

Since

gross value of output = user cost + factor cost + operating surplus,

net product may be defined in two different ways: either as

net product = gross value of output — user cost

or as

net product = factor cost + operating surplus.

User cost comprises value of materials and energy consumed in the process of production, expenditures for services of other productive enterprises, allowances for depreciation and obsolescene of real capital, and insurance of real capital. Indirect taxes should be treated as an item either of user cost or of net product according to that whether Government services are included in the classification of productive activities or not, i. e. whether we are using the concept of "universal" or of "economic" Net National Product (National Income) [1].

I.

For stochastic estimation of Net National Product one of the following three approaches based on statistical notions can be used [2]:

---

[1] See *S. Sagoroff*, Wirtschaftsstatistik: Theorie der Interpretation, Teil 1. Bern 1950, A. Francke AG. Verlag, pp. 130–149.

[2] Theoretically there is a fourth possibility of estimating total net product, namely by relating the net product to a factor of production, for instance to labor (number of employees or workers) or to capital (value of real capital, acreage of land, etc.) We would then have to deal with the quotient

$$\frac{\text{net value of output}}{\text{units of the selected factor of production used}},$$

that is to say, with the ratio of two continuous variables which do not form a proportion (a relative frequency, as in case A). However, this concept can be applied successfully only when the total net product of a sector of production (i. e., net product of Agriculture, Manufacturing, Government, etc.) is computed, and not the net National Product, since the production sectors are more homogeneous as regards the use of the factors of production than the national economy.

A. Net product is considered as a *qualitative property of the units of gross value of output.* Every unit (i. e. Dollar, Franc etc.) of gross value of output represents either net product or user cost. We wish to ascertain the total share of net product in the total gross value of output for all productive enterprises in the country, i. e. the *total quota of net product,* like a demograph who is interested in finding the percentage of males or females in the total population of an area. The magnitude we have to deal with is consequently a *relative frequency* (a proportion).

B. Net product is thought of as a *quantitative (directly measurable) property of the individual enterprises.* We try to find the absolute amount of net product per enterprise. The magnitude we obtain is an *average of absolute amounts.*

C. Instead of the absolute amount of net product the *net product quota may be envisaged as a quantitative (directly measurable) property of the individual enterprises.* From the partial quotas of net product of all enterprises we compute the arithmetic mean. The result is *an average of quotas,* identical under certain conditions with the total quota of net product (see case A).

The first approach leads to the formula

$$\overline{X}_1 = Y\left(\frac{\overline{X}}{Y}\right) = Y\frac{x}{y}, \tag{1}$$

where $\overline{X}_1$ is the estimated total net product in the parent universe, $\left(\frac{\overline{X}}{Y}\right)$ the estimated total quota of net product in the parent universe, Y is the total gross value of output in the parent universe, x the total net product in the sample, y the total gross value of output in the sample and $\frac{x}{y}$ is the total quota of net product in the sample. The subscript of $\overline{X}$ denotes the number of the estimation formula.

The second approach leads to the formula

$$\overline{X}_2 = N\overline{M}_a = NM_\alpha. \tag{2}$$

Here M_α is the arithmetic mean of the net products of individual enterprises in the sample and N the number of enterprises in the parent universe; $\overline{M}_a$ is the estimated arithmetic mean of the net products in the parent universe.

In the third case the estimation formula is

$$\overline{X}_3 = Y\overline{M}_{a/b} = YM_{\alpha/\beta} \tag{3}$$

where $M_{\alpha/\beta}$ denotes the arithmetic mean of the net product quotas of individual enterprises in the sample and $\overline{M}_{a/b}$ the estimated arithmetic mean of the net product quotas in the parent universe; Y is again the total gross value of output in the parent universe.

Thus we have:

True magnitude [1], X	Estimate, $\bar{X}$
$Y\dfrac{X}{Y}$	$Y\dfrac{x}{y}$
NM_a	NM_α
$YM_{a/b}^{(b)}$	$YM_{\alpha/\beta}$

Henceforth we shall call these three sampling methods of estimating total net product as follows:

"Total quota-Method" [Formula (1)],
"Average product-Method" [Formula (2)] and
"Average quota-Method" [Formula (3)].

II.

When the "Total quota-Method" is applied, total gross value of output is the *parent universe* and the units of gross value of output are its *elements*. In this case the individual productive enterprises are the *sampling units* and every productive enterprise must be looked upon as a "cluster of gross value of output", that is to say as a *cluster of elements*.

The variance of a sample proportion (of a sample relative frequency, of a sample total quota) is given—according to *Hansen* and *Hurwitz* [2]—by the formula

$$\sigma_P^2 = \frac{M-m}{M-1}\,\frac{PQ}{mN}\,[1+\varrho(N-1)]\,, \tag{4}$$

where

		Our symbols (case A)
M	= number of sampling units in the parent universe (or in the entire stratum)	N
m	= number of clusters sampled	n
P	= proportion of elements in the parent universe (or in the stratum) having a particular property	$\dfrac{X}{Y}$
Q	= $1-P$	$1-\dfrac{X}{Y}$
p	= proportion of elements in the sample having a particular property	$\dfrac{x}{y}$
N	= number of elements per cluster	
ϱ	= intra-class correlation between elements in the sampling unit	

[1] For explanation of the symbol $M_{a/b}^{(b)}$ see below, p. 144.

[2] See "Relative Efficiency of Various Sampling Units in Population Inquiries", Journal of the American Statistical Association, Vol. 37, March 1942, pp. 89/90.

For a finite universe

$$\varrho = \frac{\dfrac{\sum (Pi - P)^2}{M} - \dfrac{\sum Pi\,Qi}{M(N-1)}}{PQ} \tag{5}$$

where the summations are over all sampling units in the universe (in the stratum).

In formula (4) the term $\frac{PQ}{(m\,N)}$ is "the variance of a sample proportion if $m\,N$ elements are sampled independently at random, with replacement; $\frac{(M-m)}{(M-1)}$ is the finite universe correction and $1 + \varrho\,(N-1)$ is a factor measuring the contribution to the sampling variance of cluster sampling". When $N = 1$, the third term becomes unity.

Formulae (4) and (5) show that the application of the "Total quota-Method" is a difficult matter. Actually, since the *clusters are not of equal size* (gross value of output varies from enterprise to enterprise very much), other, still more complicated formulae have to be used. For this reason the "Total quota-Method" could hardly be considered for practical purposes.

III.

Proceeding to the determination of the *sampling error* we shall confine our analysis to the "Average product-Method" [Formula (2) and the "Average quota-Method" (Formula (3)].

If e denotes the "absolute sampling error", that is to say the absolute limits in which the estimate $(\overline{X})$ with certain probability deviates from the true magnitude of total net product (X), we may write

$$X = \overline{X} \pm e. \tag{6}$$

According to formula (2)

$$\overline{X}_2 \pm e_2 = NM_a \pm N e_{\overline{M}_a}. \tag{7 a}$$

Hence

$$e_2 = N e_{\overline{M}_a} \tag{8 a}$$

or, in words, the *absolute sampling error of total net product*, estimated by the "Average product-Method", is equal to the absolute sampling error of the estimated average net product of the productive enterprises multiplied by the number of the enterprises in the parent universe.

Similarly, for the "Average quota-Method", according to formula (3),

$$\overline{X}_3 \pm e_3 = YM_{a/\beta} \pm Y e_{\overline{M}_{a/b}}, \tag{7 b}$$

and

$$e_3 = Y e_{\overline{M}_{a/b}}. \tag{8 b}$$

From the general theory of statistics we know that

$$e_{\overline{M}_a} = k\,\sigma_{M_a}, \tag{9a}$$

or

$$e_{\overline{M}_{a/b}} = k\,\sigma_{M_{\alpha/\beta}} \tag{9b}$$

that is, the *absolute sampling error of the estimated average net product* (or estimated average net product quota) of the enterprises in the parent universe is equal to k—times the standard deviation of the distribution of average net product (or average net product quota) of enterprises in the sample, if $\binom{N}{n}$ samples are drawn. The coefficient k determines the probability that—so far as variation due to pure chance is concerned—the estimate will not differ from the true magnitude by more than the absolute sampling error. Thus, if we choose $K = 2$, there are 95 chances in 100 that $\overline{M}_a$ (or $\overline{M}_{a/b}$) will not differ from M_a (or $M_{a/b}$) by more than $\pm\, e_{\overline{M}_a}$ (or $e_{\overline{M}_{a/b}}$), or—to use the more fashionable terminology of the Anglo Saxon statisticians—that the estimate will lie within the "95 percent confidence limits".

The *standard deviation of the distribution of the average* net product of the enterprises in the sample, if $\binom{N}{n}$ samples are drawn from the parent universe[1], is

$$\sigma_{M_a} \sim \sqrt{\frac{N-n}{(N-1)\,n}\;\frac{\sum_{i=1}^{n}(\alpha_i - M_a)^2}{n-1}}$$

or

$$\sigma_{M_a} \sim \sigma_a \sqrt{\frac{N-n}{(N-1)\,n}} \tag{10a}$$

where σ_a is the standard deviation of the distribution of net product of enterprises in the sample, N the number of enterprises in the parent universe, n the number of enterprises and α_i the net product of the i-th enterprise in the sample.

Similarly, the *standard deviation of the distribution of the average net product quota* is

$$\sigma_{M_{\alpha/\beta}} \sim \sqrt{\frac{N-n}{(N-1)\,n}\;\frac{\sum_{i=1}^{n}\left(\frac{a_i}{\beta_i} - M_{\alpha/\beta}\right)^2}{n-1}}$$

or

$$\sigma_{M_{\alpha/\beta}} \sim \sigma_{\alpha/\beta} \sqrt{\frac{N-n}{(N-1)\,n}}\,. \tag{10b}$$

[1] Or in other terms—the standard deviation of the sample means of net products in a universe of ${}_{N}C_{n}$ samples.

Consequently the *absolute sampling error of total net product will be*

$$e_2 = Nk\sigma_a \sqrt{\frac{N-n}{(N-1)n}} \tag{11a}$$

or

$$e_3 = Yk\sigma_{a/\beta} \sqrt{\frac{N-n}{(N-1)n}} \,. \tag{11b}$$

These relations hold true under three conditions: first, that n, the number of enterprises in the sample, is sufficiently large; second, that any enterprise of the parent universe had the same chance to be chosen for the sample (random sampling); and, third, that the choosing was carried through according to the stochastic scheme "drowing of chips from a bowl without replacement" (sampling from a finite universe).

For the *relative sampling error* (ε) we have

$$\varepsilon = \frac{e}{\overline{X}} \tag{12}$$

and

$$\varepsilon_{\overline{M}_a} = \frac{e_{\overline{M}_a}}{\overline{M}_a} \tag{13a}$$

or

$$\varepsilon_{\overline{M}_{a/b}} = \frac{e_{\overline{M}_{a/b}}}{\overline{M}_{a/b}} \,. \tag{13b}$$

IV.

It must be pointed out that formula (6), namely $X = \overline{X} \pm e$ does not give an adequate expression of the relation between the true magnitude and the sampling estimate of total net product when the "Average quota-Method" is used. In the estimation formula $\overline{X}_3 = YM_{a/\beta}$ we put the mean $M_{a/\beta}$ and not the mean $M^{(\beta)}_{a/\beta}$, as it should be according to the relations in the parent universe [1], where $X = YM^{(b)}_{a/b}$.

The correct formula in this case would be

$$X = \overline{X}_3 + Y(M^{(b)}_{a/b} - M_{a/b}) \pm e_3 \,. \tag{14}$$

---

[1] $M_{a/b}$ (or $M_{a/\beta}$ for the sample) is the arithmetic mean of net product quotas, if the net product quota is considered as property of the individual enterprise, i. e. when the numbers of enterprises having equal quotas are the frequencies in the distribution of the quota.

$M^{(b)}_{a/b}$ (or $M^{(\beta)}_{a/\beta}$ for the sample) is the arithmetic mean of net product quotas, if the net product quota is considered as a property of the individual unit (1 Dollar, 1 Mark etc.) of gross value of output, i. e. when the gross values of output of enterprises having equal quotas are the frequencies in the distribution of the quota.

But the difference $M_{a/b}^{(b)} - M_{a/b}$, which refers to the parent universe, is actually not known and in practice must be left out. This is a kind of error which has nothing to do with the sampling error.

We are compelled to take $M_{\alpha/\beta}$ instead of $M_{\alpha/\beta}^{(\beta)}$ in the estimation formula (3), i. e. to refrain from weighting the average quota of net product of the enterprises in the sample with the gross value of output of the enterprises, because $\sigma_{M_{\alpha/\beta}^{(\beta)}}$, the standard deviation of the so weighted average quota, *cannot be interpreted as sampling error.*

Working with

$$\sigma_{M_{\alpha/\beta}^{(\beta)}} = \sqrt{\frac{N-n}{(N-1)\,n} \sum_{i=1}^{n} \frac{\beta}{B} \left(\frac{\alpha_i}{\beta_i} - M_{\alpha/\beta}^{(\beta)}\right)^2}$$

involves the assumption that the units of gross value of output (and not the productive enterprises) are our elements $\left(\text{as } \sum_{i=1}^{n} \beta_i = B\right)$.

But as the value units are actually grouped in enterprises, the enterprises must then be reckoned as sampling units. We encounter again the "clusters" which we had already to do with in paragraph II.

Therefore we must either regard the net product quotas as a property of individual enterprises and use the "Average quota-Method" with $M_{\alpha/\beta}$ or consider them as a property of the units of gross value of output and apply the "Total quota-Method".

V.

The choice between the "Average product-Method" and the "Average quota-Method" depends in the first place upon the nature of the available statistical data which are not provided by the sampling survey itself. The "Average product-Method" can be applied only if the number of enterprises in the parent universe (N) is known. The "Average quota-Method" requires data on the gross value of output in the parent universe (Y).

If both kinds of additional data are available, consideration must be given to the variability of the absolute amount of net product and the variability of the quota of net product in the parent universe as well as to the difference $M_{a/b}^{(b)} - M_{a/b}$.

When the difference $M_{a/b}^{(b)} - M_{a/b}$ is small, the estimation formula (3) is *more efficient* than the estimation formula (2) since the *variability of the absolute amount of net product is usually much greater then the variability of the net product quota.* The importance of variability can be seen from the following relations:

$$\varepsilon_2 = \varepsilon_{\overline{M}_a} = k\, V_a \sqrt{\frac{N-n}{(N-1)\,n}} \tag{15 a}$$

and

$$\varepsilon_3 = \varepsilon_{\overline{M}_{a/b}} = k\, V_{a/b} \sqrt{\frac{N-n}{(N-1)\,n}} \tag{15b}$$

where V_a and $V_{a/b}$ denote the *coefficients of variation* $\left(\frac{\sigma_a}{M_a} \text{ and } \frac{\sigma_{a/b}}{M_{a/b}}, \text{respectively}\right)$. The results differ in proportion to the coefficients of variation.

The difference $M_{a/b}^{(b)} - M_{a/b}$ is small when *both* the distribution of net product quota with numbers of enterprises as frequencies and the distribution of net product quota with gross values of output as frequencies are *symmetric*. Necessary condition for the symmetry of the distribution of net product quota with gross values of output as frequencies is that the gross value of output and the quota of net product of the single enterprises are not correlated. From the economic point of view this means that the quota of net product is independent of the size of the productive enterprise. Such a state of things is very probable when the sampling survey relates to Net National Product, as the parent universe embraces in this case many branches of production.

An investigation made with the help of the Central Statistical Office of Bavaria on the basis of material from the 1936 Census of Industry for Bavaria has shown the following distribution of net product quota (according to data from the sample):

Net product quota	Frequencies	
	Number of enterprises	Gross value of output (Million Reichsmark)
under 5	3	2,0
5–10	3	1,3
10–15	15	1,6
15–20	15	2,4
20–25	20	4,8
25–30	28	4,6
30–35	26	9,3
35–40	38	27,9
40–45	58	28,6
45–50	56	20,0
50–55	58	26,1
55–60	53	23,1
60–65	57	10,0
65–70	47	18,0
70–75	47	16,6
75–80	52	27,0
80–85	40	6,3
85–90	28	3,2
90–95	26	8,7
95 and over	47	6,7
	717	248,2

In this case the distribution of net product quota with gross values of output as frequencies was fairly symmetric (hence the quota of not product was factually independent of the size of the productive enterprise), while the distribution of net product quotas with numbers of enterprises as frequencies displays a positive skewness.

Against the "Average quota-Method" can be said it is subject to an uncontrolled error determined by the deviation of $M_{a/b}$ from $M^{(b)}_{a/b}$. But this method also has *a great practical advantage* in that it permits using data on total gross value of output for estimation of Net National Product. In the present state of official statistics this is important. Since the average (total) quota of net output of all enterprises over a short period of time is very stable, it is actually not necessary to undertake a sampling survey of Net National Product every year. If current statistical observation supplies yearly data on total gross value of output *it suffices to establish the total quota (average quota) of net product by a sampling survey once in three years and to apply it to the total gross value during the whole period.*

An idea of the stability of the average net product quota of all enterprises is given by the following table:

Average Net Product Quota for Agriculture

(Ascertained by the sampling method)

Agriculture Years	Switzerland	Denmark	Norway	Sweden
1937/38	0,632	0,543	0,571	0,589
1938/39	0,604	0,561	0,530	0,588
1939/40	0,614	0,548	0,536	0,612
1940/41	0,628	0,631	0,598	0,617
1941/42	0,668	0,589	0,626	0,580
1942/43	0,681	0,612	0,649	0,627
1943/44	0,674	0,617	0,678	0,602
1944/45	—	0,624	—	—
Arithmetic mean	0,643	0,591	0,598	0,602
Standard deviation	0,029	0,033	0,052	0,016
Coefficient of variation	0,04	0,06	0,09	0,03

The table is based on data of the Farm Accounting Statistics for gross value of output ("gross return") and net product ("social income") per hectare, collected according to the sampling method from many countries by the International Statistical Institute in Rome and worked out by *J. Deslarzes* [1]. The countries represented in the table are chosen for the completeness of the respective statistical series. The fact that the figures relate to agriculture only does not diminish their importance, as agriculture is that sector of production in which—owing to the influence of the weather upon physical yields and, consequently, upon gross return—the greatest yearly fluctuations of the net product quota should be expected.

[1] See the volume published in 1946 for the years from 1937/38 to 1944/45, pp. 44–46, 48–50.