ENTERPRISE ORGANIZATION AND INTERNAL INFORMATION SYSTEM : WHAT LESSONS FOR ENTERPRISE ANNUAL SURVEYS ?

Emmanuel RAULIN, INSEE
18, bd Adolphe Pinard 75675 PARIS Cedex 14 FRANCE

KEY WORDS: operational unit, statistical unit, legal unit

Enterprise annual surveys are the major element of the enterprise statistical system. Those surveys collect data for legal units on the main annual accounting data, the increase in fixed assets, the volume of work (data on employees), the breackdown of turnover by commodities. They concern most of activity sectors (4 digit level); each enterprise contributes to the sector corresponding to its main activity. Those surveys are one of the main sources for national accounts. Every year, 200 000 enterprises (legal unit) are surveyed. Data for local units are simultaneously collected : volume of work, investment, special expenses (energy, environment) for the most important establishments in manufacturing industry sector, volume of work and turnover for establishments in trade.

The rationale and form of the EAEs date back to the 1960s.

In 1992, an investigation was conducted on the desirable development of those structural surveys. Two main aims among many are the following : having on the one hand more localized data (like turnover, added value) than at the moment, on the other hand data more homogeneous by industry. This second aim is important for national accounts. At present, most cells of the sector to product transfer matrix are estimated for the added value, because many enterprises have more than one elementary activity, and contributes to more than one industry. For the same reason, it's very difficult, today, to have accurate data by administrative region.

Within the context of these investigations, 131 enterprise interviews were realized ; all those enterprises have more than one local unit. Those surveys basically concerned enterprise information system in connection with its organization : what kind of data are available, at which level within the firm ? We initially tried to see whether we could observe directly some accounting data (and particularly added value) at a local unit level. As a matter of fact, if local unit could become the basic statistical unit, then both previous problems would be solved.

Main findings of these investigations are as follows :

- most of expenditures (including investment) can be provided by the firm for each of its local units ; but, in most cases, added value is not directly available at a local unit level.

- the majority of manufacturing industries have an internal organization by operational unit (business unit) and an information system connecting with it.

- the legal unit, which is the main observation unit for the french structural statistics, doesn't seem to be always the right observation unit (especially in manufacturing industry sector). Sometimes specific units, between firm and enterprise-group or within legal unit seem to be better.

The main elements about the interview-questionnaire were as follows.

It was divided into three parts : the first one was about internal enterprise organization and its recent development (with the frame of this organization) as well as autonomy level of local or operational units, with regard to some main points : employees (remuneration and engaging), purchase, subcontracting and sales. The second one concerned the accountancy customs of the firm, of its local or operational units, beside legal duties. Finally, a third part studied, for each variable, its availability by location and where it was booked (either in the enterprise or in the local unit).

It is important to point out that most interviews involved members of the firm managing staff, as the General Secretary or the Financial Director in most cases, hardly ever the accounting Department.

Local unit : cost centre or profit centre ?

As a second point, a main lesson can be drawn : added value is available at a local unit level only in one third of cases.

Of course, this rate differs with the sector : as in the Manufacturing Industry Sector this percentage only reaches 34% of the achieved interviews, in the Trade Sector in return, added value, or more exactly working products or working costs, is available for location in 47% of cases.

Nevertheless, even when these data can be collected for each local unit, in 50% of cases, the enterprise head office is designed as the booking place.

SECTOR	VALUE ADDED AVAILABLE BY LOCAL UNIT	%
MANUFACTURING INDUSTRY	17/50	34
FOOD BEVERAGE AND TOBACCO INDUSTRIES	2/11	18
CONSTRUCTION	1/6	
TRADE	16/34	47
TRANSPORT	4/13	31
SERVICES	6/17	35
TOTAL	46/131	35

Therefore, it is possible to here conclude that, at least in France, a location can't replace a legal unit, that is to say the enterprise, for the follow-up of trading results. As a major reason, in most cases, a local unit is considered by the enterprise as a cost centre and not as a profit centre : if such is the case, costs directly bound with local unit activity are available for each location but often are not there connected with production. A local unit is not considered by the firm as the unit relevant to follow-up added value or profit.

It is significant to point out that expenditures in immaterial, as advertising, professional training, research and development, are scarcely available at a local unit level, contrary to capital expenditures for machinery or buildings. This point is all the more significant because economic studies underline the importance of intangible factors.

The minor position occupied by location in the enterprise organization seems to be a well-established movement, since the only examples we noted during those interviews, concerned local units that moved from profit centre to cost centre. Under those circumstances, it's only natural to observe that, particularly in the Trade Sector or in Close Services, trading accounts is most often available for location; it's there obvious that local units, that is to say stores or agencies, are places where profit is making up.

Operational units : more relevant units ?

Nevertheless, the major lesson of those interviews is about the importance of operational units, as establishments and departments in the enterprise organization.

Enterprise activity is really managed through those units. It becomes clear that this organization is in fact the most frequent in the Manufacturing Industry Sector. 46% of the consulted enterprises are in this situation. These operational units generally correspond to the enterprise will to the follow-up of its profitability by great family of commodities : a line of products is in fact most frequently quoted as establishment base ; in a car equipement enterprise, for example, the three settled operational units correspond to petrol engine for the first one, Diesel engine for the second one and gearbox for the third one.

Sometimes, the operational sharing can be far more discriminating and cross a line of products with specific markets (hypermarkets or specialized stores, for example). As we can see, those units can now and then meet the notion of homogeneous production unit, sometimes the one of economic activity unit. What must be stressed is the acuteness of sharings, generally greater than what requires an enterprise statistician.

If the production process for a given line of products involves several local units, then the operational units aggregate locations or parts of them : if that is the case, there is no more simple connection between local unit and operational unit.

Those operational units are then taken by enterprise managers as the relevants one for activity follow-up : main control balances, including trading results, are then established for those units. If so, immaterial investments are generally identified by establishment. Those units, settled by the enterprise head office, are free from potential legal duties. They are most often profit centres. Their accountancy integrate in some cases general costs divided according to keys peculiar to each enterprise.

SECTOR	FIRMS WITH OPERATIONAL UNITS	%	TRADING ACCOUNTS AVAILABLE BY OPRATIONAL UNIT
MANUF. INDUSTRY	23	46	18
FOOD BEV. & TOB. IND.	3		2
CONSTRUC-TION	2		1
TRADE	9	26	5
TRANSPORT	2		2
SERVICES	6	35	5
TOTAL	45	34	33

The fourth remark will concern operational units and legal units.

In many cases, it appeared that those operational units went beyond the frame of legal unit ; for 20% of the consulted enterprises, operational divisions aggregate different legal units, or at least are transversal to several legal units. What is concerned in this kind of configuration, is the important gap that can exist between the legal definition of the enterprise and the economic one.

Finally, it's necessary to stress that when those units extend beyond the limits of legal units, they often correspond to departments of a business-entity, enjoying a large independance. Within the group of companies, these departments are taken as full companies, but without legal basis ; they therefore have an entire information system.

We can then wonder whether operational units, as previously described, could constitute new statistical units.

An intermediate strategy

The direct integration of those units in a statistical system present some obstacles.

The first one consists in the great diversity of their definition from one enterprise to another : enterprise organization in establishments or departments is directly connected to strategic choices decided by the firm. Besides, the acuteness of enterprise sharing in operational units is very often too subtle with regard to statistical needs.

The second one concerns the rather confidential nature of the data : most of the consulted enterprises consider those data by operational units as strategical one from a competition point of view.

Finally, the last obstacle lies in the possible unsteadiness of such units ; do they support far better enterprise structural reorganization than legal units, or more simply, don't they highly suffer from fashion effects in matters of enterprise management ?

As a conclusion, it then seems that those operational units can't be directly considered, without any adjustment, as statistical units. An intermediate strategy can therefore be formulated.

Such a strategy must rest on the principle that we can only collect data that are controlled by the enterprise itself, for the follow-up of its activity, or data that respond to legal duties.

The search for local accounts data, directly observed, as the search for data more homogeneous with regard to activity (and particularly for the added value), must then consider enterprise internal information systems. An intermediate strategy could therefore consist in settling, with the agreement of the enterprise, statistical units that were compatible both with the enterprise internal information system, and thus with its organizational structure, and with our activity or geographic nomenclatures. Thus those units should be intermediate between operational units and legal units.

Those statistical units should afterwards be treated and updated. Such a treatement must be closely connected to the legal units register. That is all the more important because we have to use more and more administrative data (in relation to the heavy pressure to reduce the statistics burden weighing on business). The aim consists in establishing statistical units closest to both economic and enterprise organizational reality.

A first experiment will take place in France, within the framework of enterprise annual surveys, concerning only units including more than 1000 employes. The project consists in integrating gradually those new statistical units to enterprise annual surveys from 1996 onwards.

PORTUGUESE ENTERPRISE PANEL

Maria Helena Gomes, National Statistical Institute of Portugal
Av. António José de Almeida, nº 5, 1078 Lisboa Codex, Portugal

KEY WORDS: Panel, stratification, Bernoulli process, sampling errors.

1. OBJECTIVES

The Portuguese Enterprise Panel has been produced, by the National Statistical Institute of Portugal since 1990, with two main objectives. On the one hand, to supply information to the provisional National Accounts and on the other hand to supply some information about the financial variables that characterize the enterprise activities at a short time. The results concerning a certain year are published the next year, in July.

2. POPULATION

The universe of the panel is a file of enterprises created by the National Statistical Institute of Portugal, based on administrative sources. Update of the file is achieved by administrative sources as well as enterprise surveys, but not through the panel data.
The survey population is the set of enterprises with head-office at the Portuguese mainland. It excludes the enterprises with zero employees or classified as "non active". The enterprises of the regions: Azores and Madeira are excluded too, because their contribution to the variable turnover is poor.

3. SAMPLE DESIGN

3.1- Stratification

The survey population is stratified by the variables: economic activity (subdivision of the economic activity classification, two digits), juridic form (public sector, individual enterprises, others), geographical region (NUTE II level two of the geographical nomenclature: Norte, Centro, Lisboa e Vale do Tejo, Alentejo, Algarve), number of employees (three levels: 1 to 19, 20 to 499, 500 or more).

3.2- Dimension of the sample

The dimension of the sample n, is evaluated assuming that the coefficient of variation of the variable turnover is smaller than 20% for the subdivision (two digit) and smaller than 10% for the division (one digit) of the economic activity classification. In 1990, 8340 enterprises have been selected. In 1991, 9377 enterprises have been selected.

3.3- Distribution of the sample in the strata

The selection of the enterprises in the strata is done by the Bernoulli process. So the dimension of the sample, n is a random variable with binomial distribution. It's parameters are N_h, the dimension of the stratum and $f_h = n_h/N_h$, the sampling rate in the stratum. The enterprises with 500 or more employees belong to the sample with probability one, and in each stratum two enterprises or more are selected. The minimal number m_h of selected enterprises in each stratum is evaluated to a confidence level of 99%, such that $P(m_h \leq 1) \leq 0.01$. The table below shows the minimal number of the sampling dimension associated with the stratum dimension:

Table 1. Sample minimal dimension

N_h	3	4-5	6-9	10-27	28 or more
m_h	3	4	5	6	7

N_h= stratum dimension
m_h = sampling dimension of the stratum

The adition results of the number of enterprises that belong to the exhaustive strata and the minimal number of selected enterprises for the non-exhaustive strata have been calculated. The difference between the dimension of the sample and that total is distributed by the non-exhaustive strata proporcionally to $N_h\ S_h \sqrt{X_h}$. S_h is the standard deviation of the turnover for the enterprises of the stratum h and X_h is the turnover of all the enterprises of the same stratum. If the relative sampling error of the variable turnover in the strata is not smaller than 20% for the subdivision, or 10% for the division of the economic activity classification, the dimension of the sample will be increased. Proceedings will be restart. It is an iterative procedure.

3.4- Selection

The selection of the sampling units is made by the Bernoulli process.
A random number u_i with uniform distribution in [0,1] is selected and it is assigned to the enterprise i
If $u_i \leq f_h$ (sampling rate of the stratum h to which belongs to the enterprise i) the enterprise i is selected. The number u_i is retained to allow the selection of the panel, the next year.

4. QUESTIONARY

The questionary has seven items: identification and characterization of the enterprise (identification number, social designation, geographical localization, address, juridic form, main economic activity), enterprises situation (active, non active), number of employees (without salary, total), costs and losses (costs of goods and services, indirect taxes, labour costs), benefits and earnings (sales of goods and services), investments, stocks.

5.NOMENCLATURES

To codify the answers of the questionary some nomenclatures are used: Economic Activity Classification (CAE, 73), Geographical Regions Nomenclature (NUTE) Juridic Form Classification. The other codes are included in the questionary.

6. MODE OF DATA COLLECTION

The information is collected by mail, normally, in the first three months of the year. Enterprises can answer by mail or by fax. A set of instructions, a response envelope and a letter explaining the objectives of the panel are sent to the enterprises along with the questionary.

7. NON-RESPONSE

In 1990, the non-response rate was 14% and in 1991 23%. To reduce the non-response rate, two attempts are made, by mail. The two kinds of non-response: partial and total, are treated in the same way.
Partial non-response: the mean of the variable is imputed to this kind of non-response. Although this procedure reduces the estimated error, it is better than considering a null value.
Total non-response: the set of responses is the new sample dimension and the computation of the sampling errors are based on it.

The presented results concern 1990 and 1991

Tabela 2. Distribution of responses and non-responses by economic activity, 1990

Economic Activity (CAE)	Response							Non response	
	Survey	Total	%	Actives	%	Inactives	%	Total	%
Agriculture	448	1219	84%	728	60%	491	40%	229	16%
Extra.Industry	173	151	87%	133	88%	18	12%	22	13%
Trans.Industry	2 000	1766	88%	1630	92%	136	8%	234	12%
Elect. Gas. Wat.	53	44	83%	41	93%	3	7%	9	17%
Construction	287	244	85%	221	91%	23	9%	43	15%
Trade	1666	1457	87%	1353	93%	104	7%	209	13%
Transp.Comun.	209	190	91%	184	97%	6	3%	19	9%
Finan.Institu.	1312	1114	85%	962	86%	152	14%	198	15%
Services	1192	1001	84%	887	89%	114	11%	191	16%
Total	8340	7186	86%	6139	85%	1047	15%	1154	14%

Tabela 3. Distribution of responses and non-responses by economic activity, 1991

Economic Activity (CAE)	Response							Non response	
	Survey	Total	%	Actives	%	Inactives	%	Total	%
Agriculture	1547	1155	75%	769	67%	386	33%	392	25%
Extra.Industry	192	151	79%	138	91%	13	9%	41	21%
Trans.Industry	2026	1348	67%	1224	91%	124	9%	678	33%
Elect. Gas. Wat.	62	48	77%	43	90%	5	10%	14	23%
Construction	431	354	82%	312	88%	42	12%	77	18%
Trade	1807	1525	84%	1415	93%	110	7%	282	16%
Transp.Comun.	263	237	90%	229	97%	8	3%	26	10%
Finan.Institu.	1483	1213	82%	907	75%	306	25%	270	18%
Services	1422	1121	79%	995	89%	126	11%	301	21%
Total	9233	7152	77%	6032	84%	1120	16%	2081	23%

8. COMPUTATIONAL CHAIN

A friendly software was designed to compute the results of the portuguese enterprise panel. It is based on "C" language. This software is used to print the name and address of the enterprises allowing the questionary to be sent by mail.
It is used to control the responses, to register the data, to verify the data, to consult the registers, to correct or to anull them, to compute the result tables, to compute some derived variables and to guarantee the statistical confidence. The results are obtained by the software SAS.

9. STATISTICAL CONFIDENCE

Public information about one or two enterprises or any value that allows the computation of individual information, doesn't ensure the statistical confidence. These confidencial information is replaced by "..." on the results table.

10. RESULTS

There are six tables in which the panel results are presented:
. Enterprises, employees, turnover by region and juridic form.
. Turnover by employees stratum and main economic activity.
. Labour and labour costs by main economic activity.
. Costs of goods and services by main economic activity.
. Sales in the external market by main economic activity.
. Investments by main economic activity.

The presented results concern 1990 and 1991.

Tabela 4.- Enterprises, Employees and Turnover by Region and Juridic Form 1990

Region	Total	Public Sector	Individual enterprises	Others
Mainland				
Number of enterprises	263 319	291	150 908	112 120
Number of employees	2 241 985	203 395	318 474	1 720 116
Turnover	19 616 095	2 318 503	1 663 987	15 633 605
North				
Number of enterprises	96 524	37	62 113	34 374
number ofemployees	817 746	13 892	131 744	672 110
Turnover	5 770 107	68 826	618 160	5 083 120
Center				
Number of entreprises	48 273	17	32 767	15 489
Numberofemployees	307 885	2 494	65 044	240 348
Turnover	2 230 530	18 074	392 096	1 820 360
Lisboa e Vale do Tejo				
Number of enterprises	90 258	220	36 779	53 259
Number of employees	984 333	183 748	79 989	720 597
Turnover	10 652 216	2 168 083	465 543	8 018 590
Alentejo				
Number of enterprises	14 270	6	10 771	3 493
Number of employees	69 520	1 933	24 657	42 930
Turnover	496 373	45 392	121 900	329 081
Algarve				
Number of enterprises	13 994	11	8 478	5 505
Number of employees	62 500	1 328	17 042	44 131
Turnover	466 869	18 128	66 287	382 453

Tabela 5.- Enterprises, Employees and Turnover by Region and Juridic Form, 1991

Region	Total	Public Sector	Individual enterprises	Others
Mainland				
Number of enterprises	275 091	257	156 053	118 781
Number of employees	2 360 609	147 096	345 115	1 868 398
Turnover	22 756 945	1 834 315	1 902 706	19 019 924
North				
Number of enterprises	100 970	34	64 017	36 919
number ofemployees	877 974	11 182	140 566	726 226
Turnover	6 627 891	71 719	850 824	5 705 348
Center				
Number of entreprises	50 202	19	33 503	16 680
Numberofemployees	356 455	2 387	68 673	285 395
Turnover	2 459 070	17 104	360 843	2 081 123
Lisboa e Vale do Tejo				
Number of enterprises	94 025	190	38 401	55 434
Number of employees	981 581	131 389	87 484	762 707
Turnover	12 657 545	2 723 743	483 726	10 450 076
Alentejo				
Number of enterprises	15 095	5	11 285	3 805
Number of employees	70 977	875	28 425	41 677
Turnover	485 062	4 650	141 288	339 123
Algarve				
Number of enterprises	14 799	9	8 847	5 943
Number of employees	73 622	1 263	19 967	52 392
Turnover	527 377	17 099	66 024	444 253

11. SAMPLING ERRORS

The estimates are computed using the sub-population theory. The estimate of a total on the domain d is

$$\hat{X}(d) = \sum_{h=1}^{H} \frac{N_h}{rh} \sum_{i=1}^{r_h} x_{hi}(d)$$

$$x_{hi}(d) = \begin{cases} x_{hi} & , \text{if } i \in d \\ 0 & , \text{if } i \notin d \end{cases}$$

h stratum index,
H number of strata,
N_h dimension of the stratum h,
r_h number of responses in the stratum h,
$x_{hi}(d)$ value of the variable to the enterprise i of the stratum h on the domain d.

The relative sampling error is the coefficient of variation

$$CV(\hat{X}(d)) = \frac{\sqrt{Var(\hat{X}(d))}}{\hat{X}(d)} \times 100\ \%$$

where

$$Var(\hat{X}(d)) = \sum_{h=1}^{H} N^2_h \left(1 - \frac{n_h}{N_h}\right) \frac{s^2_h}{r_h}$$

and

$$s^2_h = \frac{1}{r_h - 1} \sum_{i=1}^{r_h} (x_{hi}(d) - x_h(d))^2$$

The relative sampling error has been computed by region, (level two of the geographical nomenclature), to the variables: number of employees, costs of goods and services, labour costs and investiments.

Tabela 6.- Coefficient of variation by economic activity, Mainland 1990

Economic activity	Enterprises	Employees	Labour Costs	Costs of Goods	Turnover
Total	263 319	1.73	1.83	3.68	2.82
Agriculture	20 127	5.76	5.13	18.22	13.43
Extra.Industry	1 258	8.59	8.08	16.63	7.36
Trans.Industry	48 758	2.59	2.62	3.53	2.89
Elect Gas Wat	60	0.11	0.05	0.20	0.05
Construction	25 246	8.02	8.33	10.41	11.16
Trade	124 373	3.65	4.12	5.95	5.44
Transp.Comun	8 498	4.22	3.39	16.51	6.28
Finan.Institu	10 625	3.92	15.20	23.55	15.52
Services	24 374	4.73	5.32	30.73	17.54

Tabela 7.- Coefficient of variation by economic activity, Mainland 1991

Economic activity	Enterprises	Employees	Labour Costs	Costs of Goods	Turnover
Total	275 091	1.85	3.85	5.84	2.56
Agriculture	20 463	5.50	4.96	13.86	10.04
Extra.Industry	1 321	7.87	7.35	18.04	6.96
Trans.Industry	51 385	3.42	3.50	4.13	3.68
Elect Gas Wat	66	0.09	0.04	0.20	0.05
Construction	25 211	6.67	7.70	10.66	8.89
Trade	127 632	3.04	13.72	9.74	4.82
Transp.Comun	10 064	4.33	3.51	17.32	6.45
Finan.Institu	11 541	3.46	4.60	18.27	8.80
Services	25 410	3.47	4.11	13.98	7.81

12. ESTIMATION OF THE EVOLUTION

The evaluation of the variable X is measured by

$$T_{12}(d) = \frac{X_2(d) - X_1(d)}{X_1(d)}$$

where

$T_{12}(d)$ evolution between the year 1 and the year 2

$X_2(d)$ value of the variable X in the second year

$X_1(d)$ value of the variable X in the first year

This evolution is estimated by

$$\hat{T}(d) = \frac{\hat{X}_2(d)}{\hat{X}_1(d)} - 1 = \frac{\sum_{g=1}^{G} \frac{N_g}{m_g} \sum_{i=1}^{m_g} x_{gi}(d)}{\sum_{h=1}^{H} \frac{N_h}{m_h} \sum_{i=1}^{m_h} x_{hi}(d)} - 1$$

Where

N_g dimension of the stratum g in the second year
N_h dimension of the stratum h in the first year
m_g dimension of the sample in the stratum g in the second year
m_h dimension of the sample in the stratus h in the first year
H number of strata in the first year
G number of strata in the second year

The present results concern the estimates of the evolution

Tabela 8.- Evolution of the main variables of economic activity, Mainland 1990/1991

Economic activity	Number of enterprises	Employees	Labour Costs	Costs of Goods	Turnover
Total	0.04	0.05	0.25	0.20	0.16
Agriculture	0.02	-0.08	0.03	0.16	0.09
Extra.Industry	1	0.01	0.17	0.32	0.09
Trans.Industry	2	0.10	0.27	0.18	0.25
Elect Gas Wat	3	-0.04	0.12	0.08	0,22
Construction	4	0.02	0.22	0.43	0.18
Trade	5	0.03	0.37	0.23	0.10
Transp.Comun	6	-0.04	0.14	-0.07	0.19
Finan.Institu	7	0.15	0.07	-0.07	0.00
Services	0.04	0.02	0.19	-0.19	0.03

The coefficient of variation of the evaluation is calculated by the usual formula.
The variance is calculated by the formula of the variance of a ratio.
We present results concern the estimate of the coefficient of variation of the evolution.

Tabela 9.- Coefficient of variation of evolution by economic activity, Mainland 1990/1991

Economic activity	Employees	Labour Costs	Costs of Goods	Turnover
Total	1.68	3.77	5.65	2.62
Agriculture	5.54	5.56	13.03	9.62
Extra.Industry	4.92	5.20	8.45	5.66
Trans.Industry	2.63	2.64	3.29	3.10
Elect. Gas. Wat.	0.04	0.01	0.08	0.02
Construction	7.19	5.41	10.15	8.33
Trade	3.56	13.86	9.49	5.00
Transp.Comun.	2.24	1.48	7.60	3.78
Finan.Institu.	3.89	15.37	19.45	16.45
Services	4.53	4.87	32.08	17.86

13. ACKNOWLEDGE

This panel was designed and developped by the National Statistical Institute of Portugal (Methodological and Enterprises Services) and supported by EUROSTAT with the colaboration of Mr. Pierre Lavallée from Statistics Canada. My acknowledges to Mrs. Maria Teresa Crespo, Mrs. Maria Eduarda Gois and Mrs. Maria Arminda Costa their contribution was very important to the execution of this work.

14. REFERENCES

[1] Cochran, W.G. - "Sampling Techniques", 3rd ed. Wiley, 1977.
[2] Duncan, G., Kalton, G., Kasprzyk, D., Singh, M.P. -"Panel Surveys" Wiley 1989.
[3] Grosbras, J.M. -"Methods Statistiques des Sondages", Economic, Paris, 1987.
[4] Hansen, M., Hurwitz, W., Madow, W. -"Sample Survey Methods and Theory", vol I, Wiley, 1953.
[5] Kish, L. -"Survey Sampling", Wiley 1965.
[6] Lavallée, P., Hidiroglou, M.A. - "Sur la stratification de populations asymetriques", Techniques d'enquête, vol 14, nº 1, 1958.
[7] Lavallée, P. -"Utilization de l'echantillonage aleatoire simple pour la selecction d'un panel", EUROSTAT, 1990.
[8] Lavallée, P. -"Painel de Empresas", INE 1990.
[9] Lavallée, P. -"Plains de sondage pourla selection de panels d'enterprises", EUROSTAT, 1991.
[10] Classificação das Actividades Económicas, CAE 73; INE 1973.
[11] Nomenclatura das Unidades Territoriais para fins Estatísticos, INE 1990.
[12] Sunter, A.B. -"Implicit Longitudinal Sampling from Administrative Files: A Useful Technique", Journal of Official Statistics, vol 2, 1986.
[13] Valente, C. -"Painel de empresas 90 (Análise Informática)", INE, 1991.
[14] Yates, F. -"Sampling methods for Census and Surveys" 4th ed., Charles Griffin and Company, 1981.

CREATION OF THE EUROPEAN ENTERPRISE PANEL: STRATEGIES

Seppo Laaksonen, Eurostat
Bat. Jean Monnet C5/86, L-2920 Luxembourg

The views expressed in this paper are those of the author and do not necessarily reflect the policies of Eurostat

KEY WORDS: confidentiality, data bases, longitudinal analysis, tracking, usefulness of panels

1. Introduction

The need to develop a European enterprise panel arose in Eurostat in the late 1980s. The general aim of such a panel is to produce, rapidly, provisional data on the non-agricultural sectors on the basis of a panel of enterprises. Three types of data needed to be involved: (i) micro-economic data on the enterprises with a view to a longitudinal study, (ii) macro-economic data comparable between countries, and (iii) data which meet circumstantial needs and express the attitude or reactions of the enterprises to various problems or events.

Eurostat, together with national statistical institutes, has over several years, therefore, promoted plans and projects which have been creating a base for advanced enterprise surveys and censuses. The development of the harmonized enterprise registers has been one of the most important targets, and this task has also progressed well, although much work is still left. Secondly, there is a need to focus on the less developed fields, such as surveys concerning service sectors and small and medium-sized enterprises.

The third crucial point, but not the smallest, concerns deficiencies in the methodologies used and the scarce utilization of the data material. In particular, we have observed that sophisticated panel methods are scarcely used in enterprise surveys. Essential changes of direction cannot be easily made, especially due to the hesitation of some countries. This paper focuses on the following two factors dictating such hesitation: (i) fear of increasing the response burden and the cost to enterprises and statistical offices, (ii) the scarce knowledge of panel-type surveys, their exploitation and methodologies available in these. Although these reasons are understandable to some extent, the author aims here to disprove these arguments. Section 2 considers panel designs in general terms, Section 3 presents recent ideas and aims of the planned network of European enterprise panels, and Section 4 gives a conclusion.

2. General features of panels

The panel approach in sample surveys requires that a sample be based on the panel design. This means that a file consists of identical units from both period t and period t+1 (and further t+2, ...), or that the sample from t+1 depends on the sample from t. Thus the measurement of individual changes is possible. There are several types of panel design, but the following two are most typical:

(i) A survey contains two or more distinct samples, each picked up at different points in time independently of one another; this is a rotating panel design. Attritions (deaths, nonresponse, other exits) are taken into account, but entrances have not been added to older panels.

(ii) Only one sample taken from a starting period is used, but a panel will be updated continuously regarding births and deaths and other changes so that the sampling weights of a statistical year provide scope to analyse cross-sectional population figures as well. The latter point may be difficult in long panels, since the response burden will be increased and nonresponse as well. Therefore some years later a new survey start will be necessary but some overlapping with the previous panel survey is useful.

Panel or other longitudinal analyses may, often be carried out more easily using population-level data files derived from registers or censuses.

The choice of an ideal observation unit is not trivial in any survey, but in longitudinal ones it is more problematic because of the follow-up

mechanism. For enterprise surveys it is natural to choose 'enterprise' as a key unit, and to include in the same file enough information on other levels, on lower and higher units. Typical lower units are local units and kind of activity units, which can be interpreted as members of an enterprise family' like in family surveys (the next lower level is also possible: the children of family members'). Correspondingly, the higher units are various types of enterprise groups. Estimates for enterprises can be determined directly or computed, for instance, as sums or averages of the values of the members. The enterprise composition can be changed from time to time, but no principal tracking problems appear, if census data are considered. However, a sample case may be fairly complicated if only some of the members have been included in the sample. Analogous problems can arise when estimating figures of enterprise groups.

Many special questions appear when trying to track or follow-up 'the same unit' of an original sample. Therefore the data collection will be more difficult, for example, but we pass over other questions and discuss the concept of the same unit itself. In order to understand the alternatives, let us first consider the concept of changes, for which purpose Willeboordse (1988) gives a good background. He divides changes first into two main groups: a change can be (a) a difference between two situations, or (b) an event that takes place. On the other hand, he divides the possible changes into three main classes: (i) change of characteristics (location, size class, activity), (ii) existence (entrance, exit), (iii) structure (split-off, deconcentration, take-over, merger, restructuring, change of ownership).

The information on these characteristics should be obtainable from a good enterprise register, but usually some difficulties appear. Register information is also more or less old, which is not a problem in processing a panel sample, because a data compiler can update it, but it does give rise to problems in determining qualified sampling weights.

A panel is a follow-up study, so that one has to decide at the sample design stage, which kind of follow-up mechanism will be used. There are many alternatives, as the changes above imply. Baldwin et al (1992) describe these questions for Canadian cases in which longitudinal data are taken from business registers. Several difficulties arose and much additional work was needed in creating appropriate follow-up mechanisms based on so-called *longitudinal identifiers*, since business registers were not planned for this type of use. More generally, the follow-up mechanism depends on the resources, the duration of a panel and the aims of panel analyses. It is obvious that the more thorough a follow-up mechanism is, the broader the analyses possible. On the other hand, the estimation problems will be increased if a very thorough mechanism is chosen. We shall pass over these details, here.

A proper sample-based panel thus necessarily requires a certain sampling design and a follow-up mechanism determined strictly in advance. This is not a sufficient condition: one has to take advantage of this approach and try to cope with its disadvantages. We consider both here, starting with the advantages:

(1) A panel increases appreciably the scope for monitoring data quality, also in respect of cross-sectional surveys. This alone could be reason enough for introducing some kind of panel design for all regular enterprise surveys.

(2) Although a panel design has been exploited only in cross-sectional analyses it usually improves the accuracy of the change estimates, because correlations between successive observation values are often positive.

(3) Advantage arises from the fact that it is easier to collect data from the same units since the respondents have already learned to complete a questionnaire, provided their motivation succeeded on the first occasion. On the other hand, similar errors may be repeated and a sample may be selective with respect to nonresponse.

(4) A panel can be designed so that it simultaneously gives cross-sectional and longitudinal estimates. Thus we do not lose anything by using a panel design.

(5) Due to outliers, extreme values and dramatic changes in some values over time, certain estimates may be too sensitive. Although this may be seen through standard errors, it is not enough if more robust (less sensitive) estimates are needed. A panel gives an advantage: the simplest solution for robusting estimates is to form the (weighted moving) averages of the successive values analogously to methods in time series analysis.

(6) A panel gives a certain amount of additional information via which, firstly, the effect of the unobserved (latent) time-invariant variables may

be isolated (fixed-effects models, see Vainiomäki and Laaksonen 1992, for example) and, secondly, the effect of measurement errors may to some extent be analysed.

(7) A panel gives a number of new alternatives for analyses, for example the dynamics of individual and group-level changes, enterprise demography incl. life expectancy, transition probabilities, building of advanced econometric and event-history models, and forecasting.

The worst disadvantage is the handling of data, which is methodologically more demanding than in cross-sectional analyses. It also requires the construction of additional variables such as sampling weights, change indicators and variables from several points in time. The second disadvantage is growth in nonresponse and other attrition, but this is also problematic from the view of a corresponding cross-sectional analysis. Problems of a certain type are also measurement errors which are not usually revealed in cross-sectional analyses but which are troublesome in analyses of changes (see McGuckin and Peck 1992, Vainiomäki and Laaksonen 1992, for example).

It has been argued that increased cost would be a disadvantage of a panel approach. This argument is not justifiable because the cost is the same for both high-level cross-sectional surveys and panel surveys; of course, the cost is lower for low-level cross-sectional surveys. On the other hand, it is clear that a new analysis based on panels gives rise to an additional cost, but this work is not necessary without an additional revenue or a concrete benefit.

In general, we can conclude: a panel design is the only rational alternative or a robust backbone for regular enterprise surveys. How to find a good design is not a trivial question and will be solved from case to case. Some useful new results have recently been produced by means of which both different cross-sectional surveys and panel surveys may be better co-ordinated (see Ohlsson 1992, Cotton and Hesse 1992, for example).

3. Development of a European enterprise panel network

It is possible that the above justifications for the usefulness of a panel approach are convincing enough for many statisticians and users of statistics, but they may not be sufficient for creating a European-level network of panels. It is a question also of the aims of the whole system, called EUREPAN later in this report. Lavallée (1991) presents the following general aims for such a panel network, which may be regarded as appropriate: (i) micro-economic: an individual enterprise should be distinguishable from a file; (ii) longitudinal; (iii) fresh: rapidly collected and consisting of fresh items; (iv) complex: quantitative, qualitative and strategic variables; (v) variable with respect to observation units and questions; (vi) target-oriented.

These are the demanding aims to be achieved under the same frame, thus in all countries and at a European level, but the way to attain these targets must be chosen. At a national level we should require:

- to develop enterprise registers in general and follow-up mechanisms in particular;
- to construct good sampling frames from registers or other census data, and their updated versions for estimation phases;
- to develop sampling designs in order to improve the quality of longitudinal studies;
- to co-ordinate different surveys from the view point of both cross-section and panel analyses;
- to plan data bases so that individual-level (enterprises and their sub-units) analyses from samples and censuses could be rationally organized; for example: it should form some permanent longitudinal identifiers for all units and develop computer technology so that other identifiers needed in specific research can be easily formed;
- to improve the flexibility of surveys so that it would be possible to include in them ad hoc variables and many types of variable (quantitative vs. qualitative, objective vs. subjective, historical vs. expectations for a future);
- to launch small-scale pilot surveys, which could be typically sub-surveys of regular large-scale surveys;
- to develop the target-orientation of surveys, for instance by reducing useless variables and reproducing useful ones;
- to extend advanced analysis of surveys and censuses;
- to push forward methodological development work;
- to remove confidentiality-related obstacles to the handling of individual data.

On the other hand, Eurostat together with other institutes, should be responsible for the co-ordination of the whole process and for motivating the statistical offices and other survey institutes to launch an effort for the EUREPAN. In line with this role, Eurostat should:

- develop and harmonize the important definitions, concepts, measures, etc.;
- encourage institutes to develop the documentation of previous achievements for international users and to produce new innovations in this field (support in its different forms from Eurostat is necessary);
- organize meetings in which methodologies, recommendations and applications of enterprise panels are considered;
- publish the crucial results in this field, launching a special series for this purpose, among others;
- develop international cooperation, covering also institutes in the US, Canada, Eastern Europe, Asia, etc.;
- create the basis for the rational and flexible handling of international data files.

Up to now, the statistical institutes of Europe have little experience of factual panels, although some kind of panel design is typical in most of countries. On the other hand, there are population-level data files from registers or proper censuses, which give good opportunities to construct longitudinal files afterwards or a posteriori. Such panel files may be built by needs of users, although additional work is necessary if changes in variables and in survey designs have taken place (concerning classifications, concepts, new variables), since the cross-sectional files have to be comparable from period to period and over countries (on experiences of the U.S. Census Bureau, see McGuckin 1993). It is probable that the number of comparable variables will decrease in the duration of a panel.

Panels or 'latent' panels exist in several European countries, e.g. France, the Netherlands, Sweden, Finland, Norway, Austria, Portugal and Germany, but a proper panel-type use of data files has been fairly limited, focusing on the control of quality (see point (1) in Section 2) and on a posteriori analyses (Abowd et al. 1993). A 'latent' desire to rise to a more sophisticated level is perceivable in most countries.

In addition to public institutes, there are private or partly private institutes which have experiences of enterprise panels. Typical of such surveys are business surveys (Eurostat 1989), which have a long tradition in most developed countries and have been carried out by either private or public institutes. The business surveys are intensive and give fresh information, mainly qualitative. The 'sampling' designs are varied but not very sophisticated, and their data sets are focused on large enterprises and on industrial sectors. They follow panel designs but the basic publications do not contain results of a panel type. However, the longitudinal analyses have been used in special research (e.g. Arminger 1992). We see that there are similarities between traditional business surveys and the EUREPAN, and collaboration in their development is therefore necessary.

In the ideal situation, the country-based micro-data files would be available under the same software and machine. This would offer flexible opportunities to provide all kinds of results, to construct harmonized indicators, to produce well comparable cross-country figures (standardization by models, for example) and to analyse the nature of multinational enterprises. The 'ad hoc' needs would also be easy to satisfy. This would be the only way to make full use of the data, in a sense. It is no essential disadvantage if each of the files is individually protected before the release. The perfect protection of large and unique companies is, however, impossible without destroying essential features of data.

If no individual data are transmittable by some countries to Eurostat, their handling within countries will take place. In that case, of course, Eurostat, together with participating offices, has to harmonize this work in respect to classifications, sampling designs, sampling weights, adjustments for missing values, other corrections and quality factors, among others. Correspondingly, a guideline for these purposes will be needed. Beside this, aggregated data should be obtainable for the common use. The level of aggregation and the construction of data raise problems, but data of enterprise groups, based on industries (3- or 4-digit-levels) or possibly on groups of 3-6 enterprises (called micro-aggregates in Eurostat), are not very confidential and most institutes are able to transmit this type of cross-

sectional data. In a panel more difficulties arise due to the need for a follow-up mechanism.

It is interesting to note that some private institutes have no restrictions on releasing individual-level data files for anyone who wants to buy them. For example, Bureau van Dijk has provided a CD-ROM file from enterprises of several countries, the variables of which are derived from public sources. Eurostat has tested these data files but so far they have been applied only to cross-sectional analyses. The preliminary tests have shown that the validation of files demands much work when a statistical use is concerned. These files are focused on larger companies and their use for statistical purposes is limited, although by means of post-stratification (exploiting the population-level data from official statistics) their representativeness may be improved to some extent.

4. Conclusions

The EUREPAN should be seen as a process, which implicitly already goes on in many countries, but it should be better synchronized and more cooperatively developed than recently.

The gathering together of many separate pieces will probably be the main way to make progress. One can see this in terms of a snowball approach: the small standing snowballs (latent surveys and know-how for their longitudinal handling) from different countries should get started in a good 'snow' environment; they will then automatically be increased and when, later, the balls are gathered together, a still larger and more useful snowball or a product of snowballs will be carried out.

The second (or additional for the previous one) alternative is to start a new survey, as Eurostat has done in the household sector, in which the small-scale pilot surveys are carried out in the EC countries in 1993. This type of supplementary survey is easier for making rapid progress, since the work can be started from an 'empty table,' a good prototype can be constructed from the start and the whole system can be co-ordinated well. Furthermore, the data collection can be well linked with regular surveys. However, it now seems that this approach is unrealistic in the enterprise sector due to the high starting cost.

The third alternative, close to the first one, could be to act in the same way as the Luxembourg Income Study (LIS), see Guide from June 1993. The LIS is a data bank on households and their incomes. Its files are located in the CEPS/INSTEAD in Luxembourg and are usable by researchers from countries which sponsor the LIS. At the moment 23 countries have sent their micro-level files to this bank and a number of research reports have been published. The files are mainly cross-sectional but some short panels are also available. The most essential work of the LIS has been the attempt to construct as consistently measured and symbolized variables as possible from country to country. This is no easy task due to the various systems of incomes, taxation, social benefit and surveys. The second key task has been to create a good environment for data handling so that the output requests over the world can be handled rationally (mainly using electronic mail) and the individual data remain confidential. Recently, a new micro-level data bank based on labour force surveys has been founded in the same place; it consists of files of five countries, but these are not yet very comparable.

The CEPS/INSTEAD has also started a project which aims at to create an inter-regional panel data base on firms (Gailly, 1992). Today it covers certain provinces of France, Belgium and Denmark, and the whole of Luxembourg. In addition, pilot studies are under way in three regions, two in Germany and one in the Netherlands. Some country-based panels already exist but their common use in the manner of the LIS is not yet available. Advantage should be taken of all the three CEPS data bases when developing cross-country panel data bases for enterprises and their sub-units in Europe.

The EUREPAN process will proceed slowly but surely some day covering all European countries. It will be well harmonized but give plenty of degrees of freedom for country-level applications.

References

Abowd, J., Kramarz, F. and Margolis, D. (1993). High wage workers and high wage firms: compensation policies and firm performance in France. Working document. INSEE Département de la Recherche. France.

Arminger, G. (1992). Analysis of longitudinal business test data with ordered categorical variables. *Paper prepared for the Symposium 92: Design and Analysis of Longitudinal Surveys.* Statistics Canada, November 2-4.

Baldwin, J., Dupuy, R. and Penner, W. (1992). Development of longitudinal panel data from business registers: Canadian experience. *Statistical Journal of the United Nations ECE* 9, 289-303.

Cotton, F. and Hesse C. (1992). Tirages coordonnes d'echantillons. *INSEE/France. Document de travail* E9206.

Eurostat (1989). PANSIM project. Final report. January. National Economic Development Office. Unpublished Mimeo.

Gailly, B. (1992). Inter-regional data base of panel on firms. CEPS/INSTEAD. Working Paper. Walferdange, Luxembourg.

Lavallée, P. (1991). Création d'un réseau européen par la sélection d'entreprises témoins. Projet CREUSET. Unpublished mimeo. Eurostat.

Lavallée, P. (1992). CREUSET project. Progress report. Unpublished mimeo. Eurostat.

The Luxembourg Income Study. Information guide, June 1993. Working Paper 7.

McGuckin, R. (1993). The importance of establishment data in economic research. Paper presented for the International Conference on Establishment Surveys, Buffalo, June 27-30.

McGuckin, R. and Peck, S. (1992). Manufacturing establishments reclassified into new industries: the effect of survey design rules. *Center for Economic Studies. Discussion paper* 92-14. U.S. Bureau of the Census.

Ohlsson, E. (1992). SAMU. The system for co-ordinating of samples from the business register at Statistics Sweden. *R&D Report 1992:18. Statistics Sweden.*

Vainiomäki, J. and Laaksonen, S. (1992). Inter-industry wage differences in Finland, 1975-85. Evidence from longitudinal census data. *University of Tampere. Tampere Economic Working Papers* 3/92.

Willeboordse, A. (1988). Towards a classification of changes - In search of a theoretical and practical answer to the question of birth & death. *3rd international roundtable on business survey frames.* Auckland, New Zealand.

INDUSTRIAL ENTERPRISES LONGITUDINAL PANELS IN ISRAEL: CONSTRUCTION, DEFINITIONS AND USE IN RESEARCH

Regev Haim, Central Bureau of Statistics, Jerusalem, Israel
and the Maurice Falk Institute for Economic Research in Israel
Hakirya Romema, P.O.Box 13015, Jerusalem, Israel

KEY WORDS: Industrial enterprises statistics, longitudinal, panel

1. Introduction

Statistical Bureaus collect large sets of data on a current basis (monthly and annually) on the activity of industrial establishments (i.e., production, employment, wages, investment, etc.) in their respective jurisdictions. Sample-based surveys of industrial firms are carried out for the principle purpose of providing policy makers with indicators of current trends in this sector, on its structure and on input-output relationships.

These data, available in the files of statistical bureaus and covering long periods of time, also have a potential as data sets for sophisticated economic research and analysis. Cumulative data covering individual firms ("panels") over an extended period of time offer researchers a very important new dimension — time change — which can enhance industrial structural research.

The methodology for the creation of such panel data sets and their use in econometric research has been developed in the past decade. This has opened new paths for a better understanding of the factors underlying productivity changes, profitability, failures and closures, etc.

A number of statistical bureaus have already begun arranging these data sets into panels, as is the case at the Central Bureau of Statistics in Israel. Subsequently, these panels have been made available to researchers outside the statistical bureaus, sometimes as joint ventures between bureau and university researchers. However, the creation of such panels pose a number of problems and difficulties, such as matching data from various sources for the same firm, creating the appropriate variables, filling gaps in the data, etc.

The experience of Israel in creating such a panel data set is described in this paper, as well as some of the research performed to date on this data set.

2. Panels and their use

A Longitudinal or Panel data set is defined as one that follows a given unit over time. This means that the unit of observation (the firm) varies in two dimensions. In our case we have a sample of industrial firms observed at several points in time (to use Mundlak's (1968) terminology, in panels we are pooling time-series and cross-sections). (1)

That is why a panel data set enables one to follow the development of a population of firms , controlling for time effects or, alternatively, to analyze the time effects, controlling for changes in the structure or characteristics of the firms, and analyzing the interaction between the two.

Longitudinal data enable us to analyze balanced and unbalanced panels. Balanced Panels are panels in which for each observation we have data for the whole period. Unbalanced Panels are panels where some firms were observed only in some sub-periods. For example, a panel which includes firms that operated during each year is a balanced panel. A panel which includes firms that were opened or closed during the period is an unbalanced panel.

Panel data can be used to estimate cost and production functions, labor and total productivity at the firm level, firm turnover and survival rates, life cycles, the job generation and destruction process, and many other related topics. They provide a rich environment to use complex econometric methods. Firm panels are usually confidential and researchers do not have easy access to them. That is why panels are studied jointly by Bureaus of Statistics and Research Institutes. In Israel, the panels are analyzed by a group of researchers from the Central Bureau of Statistics, the Maurice Falk Institute for Economic Research, The Bank of Israel's Research Department, NBER and Toronto University. In the USA we can find a parallel network of researchers.

3. Panel Construction and Sources of Data

The basis of our panel are the data collected regularly from samples of some 2500 industrial firms which are used to compile the monthly and annual industrial statistics. From these firms the following data are collected:

Data on sales, labor input, and cost of labor and commodities produced are collected on a monthly basis.

Industrial Surveys are now conducted annually (although in the past this was not always the case).

In these surveys data on firms' characteristics (branch, sector, ownership, locality, etc.), income, expenditures, labor cost, inventory and investments are collected.

R&D surveys are carried out on an annual basis.The survey covers R&D expenditures and technically skilled labor.

From time to time special topics have been surveyed using the full sample or a subsample of firms. The topics covered have been capital surveys (in 1968 and 1982), (2) (3) and skilled labor (a full survey was carried out in 1988 (4) and for part of the firms data was collected from administrative files in 1979 and 1984).

Our sample is drawn from the register of firms of the National Insurance Institute, which covers all employees in Israel. The register is updated regularly and new openings are included in the sample during the course of the year. Closed firms are excluded from the sample when the Bureau's enumerator reports that a firm has cease to exist and their file at the National Insurance register is closed, or when they change their activity and move to another industry.

The sample is changed once or twice per decade and adjusted for under-coverage from time to time. The previous sample was conducted in 1979 and used 'till the end of 1989, with a major adjustment in 1985. The coverage of firms with 75 or more employees is essentially complete. Each firm included in the sample is given an I.D. number which is changed only when the sample is changed. Because data are collected from the firm on a monthly basis changes in the activity or the organization of each firm is recorded and routinely checked.

The first step in the building of our panel was to match the firms in the industrial surveys carried out from 1979 to 1988 and to prepare a file with all the firms reported in at least one of the 8 industrial surveys carried out during this period. Using the records on the changes in the sample we were able to construct a variable for the status of each firm in each industrial survey if it continued as the same firm, as a new firm, as a closed firm or a firm which was included in the file against under-coverage.

The next stage was to define the variables needed to describe the development of firms so as to measure their economic performance and to estimate various economic models such as production and cost functions. The variables were calculated from the above mentioned sources of data and matched to the basic file. Missing values were estimated where necessary.

Finally, the various models were estimated, analyzed and studied carefully and where necessary changes and improvements in the various variables were incorporated. This last stage was extremely time consuming.

4. Price Adjustments

The original data from each Industrial Survey were adjusted for price changes during the year. This mainly influenced the change in inventories, by excluding inflationary profits.

Because of the high inflation rate in Israel, especially up to 1985, we adjusted the data for price changes in 2 stages using a general price index and (relative) specific price indices.

General price adjustment

Using the Consumer Price Index, the data were calculated (adjusted) to June 1990 and prices converted to USA$.

Specific price adjustment

Production and intermediate product figures were also calculated in constant prices using specific price indexes which were constructed at the 3 digit economic classification (about 100 sub-branches). For production figures we had the local sales production index and export index. They were weighted at the firm level by the relative shares of local sales and export. The price indices for intermediate products were constructed, using industrial local sales and import price indices and the 1982 input-output coefficients as weights, for 186 sub-branches. The overall index for the intermediate input price is a weighted index of local and import prices weighted by the corresponding 1982/83 weights from the basic input-output table, aggregated for the 100 industrial sub-branches.

5. Topics and Variables

A. Firms' characteristics. From the Industrial Survey we can attach to each firm a code or dummy variable for its economic branch, sector and legal ownership (private, corporate, union (histadrut) owned, kibbutz or government), locality and size.

B. Life Cycle. Establishment year, closing year, and the firm's age.

C. Labor Inputs. The measure of labor input is the 'Person Year', which is calculated as yearly hours worked divided by the potential of hours per year. The data comes from the monthly surveys.

D. Production Account at '1990 general prices'. This account covers the value of production and the payments for inputs and means of production, i.e. the cost of labor .

The components of the Production Account are: Production (value of): Sales + Inventory Change of finished and semi-finished products.

Intermediates costs: Materials Used (after inventory adjustments) plus General Expenses such as: Municipality taxes, insurance, accounting, communication, etc.

Labor Costs: Wages plus additional costs such as pensions National Insurance contributions, travel costs, meals , and clothes.

Value Added: Production minus Intermediates.

Gross Margin: Value Added minus Labor Costs.

This account enables one to calculate profitability, the factor shares needed for the TFP (total factor product) calculation and some of the variables used in the cost function.

An illustration of the production account can be found in Table 2.

Production Account at Fixed Prices. This is the framework used in the estimation of production functions, the 'quantities' of intermediate materials used, labor, and capital. All data were deflated by a complex series of relative indices which were calculated at the 3 digit industry level using local sales prices, export and import prices, and appropriate weights (as described in Section 4).

F. Capital. Modern economic production theory distinguishes between 3 types of Capital: Fixed Capital, Human Capital and Knowledge. Our experience is that these variables are the most complicated and crudest variables.

Fixed Capital Services is calculated as depreciation plus interest (5%) on the net capital of Buildings, Equipment and Cars that existed at the beginning of a period. Asset life assumed: 33 years for buildings, 14 years for equipment, and 8 years for cars. The variable was built from the 1968 and 1982 fixed capital surveys as benchmarks, and used 'annual investments' as indicators of change over time, using the perpetual inventory method.

Human Capital was calculated as an index of the quality of labor. The index equals the weighted values of engineers (1) and technicians (0.75) plus Total Person Years divided by Total Person Years.

Knowledge capital stock was estimated by R&D Capital Services. First we estimated the Capital R&D Stock as the cumulative R&D expenditures over 7 years. R&D Capital Services was then calculated using the same formula as Fixed Capital Services, i.e., depreciation plus 5% interest on net capital R&D. Knowledge life is assumed at 7 years.

For these topics we did not have all the data for all of the firms, and so used some methods of imputation. A full discussion of how these variables were constructed is the subject of a separate paper (Regev, 1993).(5)

6. Time Coverage: Periods and Consistency of Panels

Each Industrial Survey presents the picture of the current year. There are usually some changes in the sample over the years, due to changes in the economic classification of firms, under-coverage or other reasons.

The panels are organized according to the periods when new samples were introduced. The three periods are:

1. 1979-1988 — includes data from 8 Industrial Surveys
2. 1970-1977 — includes data from 6 Industrial Surveys.
3. 1958-1967 — includes part of the firms that operated in those years.

An effort was made to construct consistent panels for each of the 3 following sub-periods: 1979-1982, 1982-1985, and 1985-1988.

Each Survey covers around 2000 firms.

7. Organization of Research

One of the main incentives to invest the huge effort needed to construct these longitudinal panels was to enable the use of these data by researchers from universities and other research institutions. A group of researchers from the Central Bureau of Statistics, the Maurice Falk Institute for Economic Research in Israel, the Bank of Israel, the NBER and Toronto University have been exploring for the past year this unique set of data. In order to protect the confidentiality of the data, access to the data bank is limited to authorized researchers, and even then only via Bureau researchers. Where possible, the Bureau encourages joint research between Bureau and outside researchers.

8. Studies Completed

The following studies are based on this panel of data (as of March 1992):

High Tech and Productivity

In this study an index of high-tech was developed, based on Fixed Capital intensity, R&D activity and the vintage of fixed capital. For 1982, the firms were classified by high tech level and the connection between high technology and productivity was analyzed. Results were published in (6).

Firm Turnover and Productivity

This is the central study of the 1979-1988 panel. It traces the development of the population of firms during the period, the turnover of firms, the growth of output and productivity and analyses their correlates. A preliminary report was published in (7).

See an illustration of this kind of analysis in Table 2.

Production and Cost Function

In this study, some basic economic models were used to study the economic performance of industrial firms in 1979–1983. The models used were: a trans-log production function, a cost function, and a wage function. Some econometric problems such as sample selectivity, serial correlations due to unobserved firm effects and endogeneity are addressed. The study was published in (8).

Sex, Wages and Productivity

Our data for 1988 were used to estimate marginal productivity and wage differentials associated with the percentage of women in each firm. Hypothesis that lower wages paid to women reflect lower marginal productivity was tested. Results were published in (9).

Research in Progress:

Regev, H., The Survival of Industrial Firms in Israel

Gronau R., Regev, H., The Job Creation and Destruction Process in Israeli Industry

Regev, H., Bar-Eliezer, S., Concentration and Penetration in Israeli Manufacturing 1977-1988

Griliches, G., Regev, H., R&D and Productivity

Bregman, A., Fuss, M., Capital Subsidy

BIBLIOGRAPHY

(1) Mundlak, Y., "On the Pooling of Time Series and Cross Sectional Data", Econometrica, 46. 1978, pp. 69-86.

(2) "Survey of Fixed Capital Stock in Industry 1968", Special Series No. 413. Central Bureau of Statistics, Jerusalem, 1982.

(3) "Survey of Fixed Capital Stock in Industry 1.1.82", Special Series No. 785. Central Bureau of Statistics, Jerusalem, 1986.

(4) "Industry and Crafts Survey 1988, Part 1", Special Series No. 898. Central Bureau of Statistics, Jerusalem, 1991.

(5) Regev, H., "Estimating Capital Services in an Industrial Panel" (in preparation 1993).

(6) Bregman, A., Fuss, M., Regev, H. "High Tech and Productivity: Evidence from Israeli Industrial Firms", European Economic Review 35 (1991).

(7) Griliches, Z., Regev, H., "Firm Turnover and Productivity in Israeli Industry", 1979-1988, Working Paper No. 4059, National Bureau of Economic Research, April 1992.

(8) Bregman, A., Fuss, M., Regev, H., "The Production and Cost Structure of Israeli Industry: Evidence from Individual Firm Data". Working Paper No.4059, National Bureau of Economic Research, April 1992.

(9) Judith K. Hellerstein and David Neumark, "Sex, Wages and Productivity: An Empirical Analysis of Israeli Firm-Level Data", Discussion Paper No. 93.01, The Maurice Falk Institute for Economic Research in Israel, January 1993.

TABLE 1

Firms, Person Years Production and Value Added by Period and Turnover

Period	Total Begin	Total End	Continuing Begin	Continuing End	Replaced Begin	Replaced End
			Firms in Panel			
1979-1982	1951	1893	1646	1646	305	247
1982-1985	1881	1854	1616	1616	265	238
1985-1988	1986	2003	1759	1759	227	244
			Firms in Population			
1979-1982	6206	6239	4619	4619	1587	1620
1982-1985	6183	6391	4779	4779	1404	1612
1985-1988	6473	7055	5284	5284	1189	1771
			Person Years			
1979-1982	286695	279950	257387	256970	29308	22980
1982-1985	277759	291127	256792	273395	20967	17732
1985-1988	299381	288279	283043	269285	16338	18994

TABLE 2

Production Account, 1979-1990 (ths 1990 $)

	79/80	82/83	85/86	88	90
1. **Production**	61.19	69.30	73.74	77.82	86.16
2. **Intermediate**	43.71	46.39	49.55	49.41	57.62
3. **Value Added**	17.48	22.91	24.19	28.41	28.54
4. **Labor Costs**	15.67	19.54	18.25	21.21	21.41
5. **Gross Margin**	1.81	3.37	5.94	7.20	7.13

ILLUSTRATION A

Capital Variables, 1979-1988 (1990 th's $ per person yr)

		79/80	82/83	85/86	88
1.	Fixed Capital	4.20	5.58	6.15	7.64
2.	R&D	0.43	0.51	0.71	0.71
3.	Labor Quality Index	1.07	1.09	1.11	1.11

ILLUSTRATION B Production Functions

Dependent Variable: Production Per Person-Year

Regressions	Details	
Observations	7741	
R-Square	0.863	
Root MSE	0.300	
	Coeffs.	**T-Stat.**
Intercept	1.445	62.399
Intermediates	0.688	156.471
Capital Services	0.058	10.912
R&D Variable		
R&D	0.026	4.682
No R&D	-0.019	-1.403
Quality of Labor	0.406	6.726
Scale (Ref = 50 - 99 workers)		
5 - 49 Employees	0.003	0.351
100 - 299 Employees	-0.014	-1.225
300 + Employees	-0.012	-0.765
Mobility (Ref = Stayers)		
Closed 1979-82	-0.100	-4.931
Closed 1983-85	-0.058	-3.665
Closed 1986-88	-0.092	-6.373
Sector (Ref = Private)		
Reg. Stock Market	-0.035	-2.031
Histadrut	0.029	2.090
KIbbutz	0.042	3.107
Public Sector	0.061	2.134
Others	0.109	4.142
Branch (Ref = Electronics)		
Food	-0.035	-2.315
Textile	-0.037	-2.671
Printing, Paper	-0.051	-3.390
Wood, Mineral	-0.026	-1.661
Chemical, Plastic	-0.012	-0.883
Metal, Machinery	0.035	2.660

Note: The regression included variables for lifecycle and year dummies which are not shown due to a lack of space.