

HELLENIC STATISTICAL AUTHORITY

INTRODUCTION OF A NEW PRODUCTION SYSTEM FOR THE COMPILATION OF BUSINESS STATISTICS

Vassiliki Benaki

FIFTH INTERNATIONAL WORKSHOP ON BUSINESS DATA COLLECTION METHODOLOGY STATISTICS PORTUGAL - LISBON 19-21/9/2018 Introduction of a new production system for the compilation of Business Statistics

Business Statistics are acknowledged as a major component for policy making .

The current economic circumstances create an urgent need for delivering more comprehensive data which in turn increase the response burden of enterprises and the cost of producing such statistics, due to general resources constraints.

Traditional survey based method for the compilation of Business Statistics

The applied method for the compilation of business statistics in order to collect information for the Short Term BS, country level BS, Regional BS & statistics on International Activities up to 2014 in Greece, was based on large scale sample surveys to enterprises of various economic activities such as manufacturing, construction, trade and services



Traditional survey based method for the compilation of Business Statistics

The problems encountered for the compilation of Business Statistics were:

- Low response rate impacting the quality of the statistical output in the domain of business statistics
- Delays in the production of statistics
- Burden on the enterprises for providing the relevant data
- Requirements for significant financial and human resources due to the large number of enterprises in the samples

NEW SYSTEM FOR COMPILING BUSINESS STATISTICS

Use of Administrative Data for Business Statistics

In the context of improving the quality of business statistics and addressing all the above mentioned issues Hellenic Statistical Authority (ELSTAT) explored the use of alternative sources, besides surveys, for the compilation of business statistics. To this end it was decided the use of administrative - mainly tax - data for the production of the main statistical characteristics of businesses in Greece.



NEW SYSTEM FOR COMPILING BUSINESS STATISTICS

WHY ?

to use administrative data

Administrative data are data kept either in public registers and files or other administrative data sources and bodies of the public sector, in printed, electronic or other form.

Use of Administrative Data for Business Statistics

PROS

- Full coverage of the enterprises' population
- Overall improvement of the statistical output
- Reduction of the enterprises' administrative burden
- Significant reduction of the cost for the production of the statistical results
- Compilation of new statistical indicators in the domain of business statistics in order to meet the growing users' needs

Use of Administrative Data for Business Statistics

CONS

- Admin data are collected for non-statistical purposes
- Administrative authorities control data collection and processing
- Changes in the regulatory framework may lead to breaks in data series
- Restrictions to access due to technical and confidentiality issues

NEW SYSTEM FOR COMPILING BUSINESS STATISTICS

Strategy to get access to administrative data

- 1. Legal and Institutional Framework
- 2. Analysis of the structure of the administrative data
- **3**. Regulation of ELSTAT for the use of data from administrative sources for the compilation of business statistics
- 4. Security Policy and development of a secure IT environment

Legal and Institutional Framework

1. EU Legal Framework

- Regulation 223/2009 of the European Parliament and of the Council as amended by Regulation 759/2015
- Regulations 177/2008, 295/2008,696/1993,1165/98 716/2007 of the European Parliament and of the Council
- 2. National Legal Framework
 - Greek Statistical Law 3832/2010 as amended and in force
 - Greek Law 4174/2013 as amended by Law 4364/2015 providing ELSTAT with access to tax data

Legal and Institutional Framework

3. Institutional Framework

- Memorandum of cooperation between the Hellenic Statistical Authority (ELSTAT), the Independent Authority of Public Revenue (IAPR), the General Secretariat for Information Systems (GSIS) and the Social Insurance Institution (IKA) on their responsibilities as regards the exchange of statistical information for the updating of Statistical Business Register, and the satisfaction of the needs of Structural Business Surveys and of Greek National Accounts
- Agreement of June 2016 between Hellenic Statistical Authority (ELSTAT) and the Independent Authority of Public Revenue (IAPR) to ensure the future regular transmission of the tax data as described in the relevant annex to the agreement

The new system for compiling BS was initially decided to be implemented for the compilation of Structural Business Statistics (SBS)

In this context an analysis of the structure of the administrative data was carried out at the first stage

Aim of SBS

- To provide an estimation: on the structure and development of enterprises activities, the production factors use, the performance and competitiveness of enterprises, the policy implemented by enterprises, the special characteristics and special distribution of the enterprises activities
- To be used as input for the compilation of GDP (from the side of production, expenditure and income) and of regional data on GDP, Gross Value Added, Employment and Investment

Variables to be compiled:

- Business demographic variables i.e. number of enterprises, local units
- Input related variables
 - Labor input variables i.e. number of persons employed, hours worked
 - Cost of input variables i.e. total purchases of goods and services and personnel costs
 - Capital input variables i.e. total investments in tangible goods
- Output related variables i.e. turnover and production value
- Variables related to specific expenses categories i.e. payments for agency workers and long-term rental and operational leasing

- Identification of the statistical variables that could be derived from the admin data
- Analysis of the administrative data in terms of their content
- Comparative analysis between statistical characteristics and characteristics from administrative sources in order to identify differences in definitions

The results are grouped into 3 categories:

- characteristics completely identical

- characteristics partially identical

- characteristics for which there is no available information
- Identification of missing variables
- Compilation of a transition table from administrative to statistical variables

Transitional table from administrative to statistical variables

Inde Kuðsnög	Περιγραφή Μεταβλητής	EYNAPTHEN ANTIETOIXHENE	ПНГН ПАНРОФОРІАХ	με προηγούμενα έτη
1 11110	Αριθμός επιχειρήσεων		Με άλλο τρόπο	10x8
2 11 21 0	Αριθμός τοπικών μονάδων	E3 061 + E3 064 + E3 067 + E3 070 + E3 073 + 1	8	DX3
3 11 31 0	Αριθμός μονάδων οικονομικής δραστηριότητας		Με άλλο τρόπο	DX8
4mm 12 11 0	Κύκλος εργασιών Γ Κατηγορία	E3_459 + E3_488 + E3_863 + E3_864 + E3_865 + E3_86t	8	EXC
48a	Κύκλος εργασιών Β΄ Κατηγορία	E3 540 + E3 547 + E3 283	8	1000
466	Εναλλακτική Β' Κατηγορία	E3 344+E3 551	0	Day
50 12120	Alia magnumbe	585 12110 + 13213 +13211 - 13120 + F3 471 + F3 465* + F3 467	585 + F3	Dat
5.5	Farshammen C Kermannia	585 12110 * 13218 *13211 - 13120 + Ami(mun) 28 + 74 08 + 74 98 + 71	SRS + Inclines	
6 12130	Anothing are a finite a vision of a supply more unampilitant	585 18160 - 18120 * 18211	CRC	Dat
7 17150		CBS 12110 + 10210 + F3 471 + F3 465 + F3 467 - 585 1011/	585 + 73	Dout.
/ 12150	the second s			in the second se
8 121/0	AKEREADING ALTOUPYING TAROVEOUR	365_12190 - 13310	202	00
9a 13110	Συνολικές αγορές αγαθών και υπηρέσων Γ Κατηγορία	E3_421+E3_422+E3_423+E3_442+E3_445+E3_446+E3_451+E3_456+E3_456	6	Dia
		E3_251+E3_252+E3_516+E3_517+E3_518+E3_525+E3_526+E3_527+E3_528+		
90	Συνολικές αγορές αγαθών και υπηρέσιων Β΄ Κατηγορία	E1_529 + E1_530 + E1_531 + E1_532 + E1_533 + E1_534 + E1_535 + E1_536	8	
10a 13 12 0	Αγορές αγαθών και υπηρεσιών που αγοράστηκαν με σκοπό τη μεταπώληση στην κατάσταση στην οποία παραλήφθηκαν Γ	E3_401 + E3_402 + E3_403	8	Cox1
105	Αγορές αγαθών και υπηρεσιών που αγοράστηκαν με σκοπό τη μεταπώληση στην κατάσταση στην οποία παραλήφθηκαν δ'	E3_231 + E3_232	8	
11 13 13 1	Πληρωμές για εργαζομένους από γραφεία διαμεσολάβησης	and the second	-	DXI
12a 13210	Μεταβολές αποθεμάτων αγαθών και υπηρεσιών Γ΄ Κατηγορία	E3_450 - E3_449	8	1000
125	Μεταβολές αποθεμάτων αγαθών και υπηρεσιών Β΄ Κατηγορία	E3_521 - E3_520	63	
	Μεταβολές αποθεμάτων αγαθών και υπηρεσιών που αγοράστηκαν με σκοπό τη μεταπώληση στην ίδια κατάσταση στην οποία			
130 13211	παραλήφθηκαν Γ΄ Κατηγορία	E3 427 - E3 426	13	DX8
	Μεταβολές αποθεμάτων αναθών και υποσσιών που ανοσάστηκαν με σκοπό τη μεταπώληση στην ίδια κατάσταση στην οποία			
135	nasaladinay B' Kernyana	F3 257 - F3 256		
140 13213	Μεταδολά αποθειώται τελιούς ποσλάσται μαι παρίδισται υπό ματοπεριά από του ίδια το μομάδο Γ΄ Κατοπορίε	F1 430 + F1 431 + F1 436 - F1 420 - F3 432 - F1 435	-	Dat
148		F3 360 + F3 365 - F3 350 - F3 367	ä	
15- 13 31.0	King and State and St	F1 (19		2001
154				0.44
100 13330	norme, sporteritario e katrivopia			in an
10 13 52 0	Invitation Kas rystoppustika	100/ 0100 00.05 + E1_E popular = ETAA 2	REOLUTIO + ET	0.0
100	eventeenteen	παγ/σμος σύτου+ πογ/σμος σύτοι +πογ/σμος σύτος σύτος σύτος	Macuria	in an
17a 13 55 0	ασπαντς κοινωνικής ασφαλισής Γ.Κατηγορία	Λογ/σμος 60.03 + Λογ/σμος 60.04	Ισοζυγιο	U.G
17ß	δαστάνες κοινωνικής ασφάλισης Β Κατηγορία	585 13310 - 13320	585	
15 13411	Πληρωμές για μακροπρόθεσμη μίσθωση και λειτουργική χρηματοδοτική μίσθωση αγαθών	Λογ/σμός 62.04.10 + Λογ/σμός 62.04.20 + Λογ/σμός 62.04.25	kaoline	Exc)
19a 15 11 0	Ακαθάριστες επενδύσεις σε υλικά αγαθά Γ Κατηγορία	E3_878 + E3_879 + E3_880+ E3_471	63	Dist
195	Ακαθάριστες επενδύσεις σε υλικά αγαθά 8΄ Κατηγορία	E3_611 + E3_612	B	
20 15 12 0	Ακαθάριστες επενδύσεις σε γη		-	1D00
21 15 13 0	Ακαθάριστες επενδύσεις σε υπάρχοντα κτήρια και άλλες εγκαταστάσεις		-	DXX
22 15 14 0	Ακαθάριστες επενδύσεις σε ανεγέρσεις και μετατροπές κτηρίων		-	Dxt
23 15 15 0	Ακαθάριστες επενδύσεις σε μηχανολογικό εξοπλισμό		-	Dxa
24 15 21 0	Πωλήσεις υλικών επενδυπικών αγαθών	E3 195 + E3 870 + E3 871 + 872 + E5 873	8	1000
	Ακαθάριστες επενδύσεις σε εκγωρήσεις δικαιωμάτων, διπλώματα ευρεσιτεγγίας, άδειες λειτουργίας, εμπορικές ονομασίες και			
25 15 42 0	durach Sequiliants			DXI
26 15441	Franklinger og mynder kommunde	Ammanautri 16 17 00	len linn n	Dat
27 16110		CR5 16120 + CR5 16180	CRC	Dat
28 16120	A second s		202	Date:
20 10120		F1 045 - F1 045		Dest.
	- Charles - Char		23	~~~
	a Evennesturi i Netrijopha	001-10-005	001-10	
	a conversion i vezzitaten	cs/Tsbahhd eTrodyd 1	e,	
30 16140	ναιώνος πιαριωτικό με ταρογγάτα μεγίδρονς αμετρίζουναυς			0.0
31 16150	Αριθμος των δεδουλευμενων ωρών εργασίας από τους μισθωτούς		-	DX0
32 17 32 0	Αριθμός καταστημάτων λιανικού εμπορίου	E3_061+1	63	OKI
33 17 33 1	Επιφάνεια καταστημάτων λιανικής πώλησης	*	-	Dixia
34 18 10 0	Κύκλος εργασιών από γεωργικές, δασοκομικές, αλιευτικές και βιομηχανικές δραστηριότητες	585_18120 + €3_635	585 + 83	DXI

Pilot study

- Comparison between the survey data and the corresponding administrative data
 - Calculation of the SBS variables on the basis of the transitional table
 - Comparison between the SBS variables derived from the tax data and the corresponding survey variables, by economic sector (1, 2, 3 and 4digit level of NACE Rev.2)
 - Comparison between the values of the SBS variables and the corresponding values from the administrative sources at micro level
 - Analysis and documentation of the discrepancies
 - Assessment of the new system



Results of the first comparative analysis

A comparison was made on a basis of 12,426 common enterprises identified both in SBS sample survey data and tax data for the reference year 2014

Variables	% of enterprises with zero difference		
Turnover	83%		
Personnel Cost	89%		
Changes in stocks of goods and services	78%		
Gross investment in tangible goods	69%		

NB: The above figures referred to the exercise performed before tax data cleaning



Results of the first comparative analysis

	Turnover	Personnel costs	Gross investments in tangible goods	Total purchases of goods and services	Changes in stock
Number of enterprises with zero difference	10,348	11,103	8,539	7,580	9,637
Number of enterprises with non-zero difference	2,078	1,323	3,887	4,846	2,789
Total difference in millions Euro (SBS-tax data)	-184.2	8.5	276.2	240.9	354.5



An internal document describing all procedures for the use of admin data was drafted based on the GSBPM which is structured in 4 stages.

STAGE A: COLLECTION OF TAX DATA

- Production phase
- Transmission phase
- Consumption phase

STAGE B: DATA PROCESSING

- Data editing for updating the Business Register (BR)
- Completeness checks and corrections
- Content checks and corrections

STAGE C: PRODUCTION

- Update of Business Register (BR)
- Compilation of SBS variables
- Estimation of missing variables

STAGE D: INTEGRATION IN THE PRODUCTION DATA BASE Integration of data from different data files

STAGE A: COLLECTION OF TAX DATA

The first stage refers to the transmission of all files from the Tax Administration to ELSTAT and their integration in a secure database. At this stage, data are encrypted and transformed to a form that can be read and processed by users (ELSTAT's staff). The received data are checked in terms of quality and completeness and follow "data cleaning", "standardization" and establishment of relations between the files (data controls of first stage).

This stage includes three phases:

PRODUCTION PHASE

The data flow is directed from the Tax Authorities to ELSTAT

STAGE A: COLLECTION OF TAX DATA

TRANSMISSION PHASE

Secure transmission of data via FTP (File Transfer Protocol) Server with SSH (Secure Shell) support, administrated by ELSTAT

CONSUMPTION PHASE

- Data transfer to another server connected to an isolated network protected by Firewall

- Verification of digital signature of data
- Saving of the received encrypted data
- Decoding of data

SECURITY POLICY

- Use of semi-automated data transmission method via File Transfer Protocol (FTP) Server with Secure Shell support and access only within the secure public administration network, SYZEFXIS
- Set up of the Firewall of ELSTAT with access only to authorized persons of the GSIS
- Provision of access to authorized persons for the transmission and the reception of the data from ELSTAT and GSIS

Regulation of ELSTAT on the Use of Administrative Data SECURITY POLICY

- Encryption, signing, decoding and verification of transmitted files via the software GNU Privacy Guard
- Access of authorized users to the FTP Server
- Maintenance of complete and accurate list (log) of each action and access of the authorized users to the Server
- Data transfer to the intermediate database
- Creation of backup copies of the intermediate database
- Classified access of users to the intermediate database for data processing

Secure Database Architecture



Secure Database Architecture



STAGE B: DATA PROCESSING

PHASE A: Data Editing for error detection in order to update BR

- Technical checks
- Logical-Mathematical checks
- Quality controls

QUALITY CONTROLS

a. Evident error controls at enterprise level and/or aggregated activity sector by employment level

3(a) Systematic error detection3(b) Outlier detection

Checks are based on:

Number of persons employed Number of Establishments Wages and salaries Number of persons employed Turnover Number of persons employed

(MAD method, etc)

3(c) Historical controls (year-to year)



QUALITY CONTROLS

b. Selective Editing for the detection of Influential errors

Method: Contamination Models Implementation- SeleMix package in R programming language

Enterprises with "influential" errors (errors which significantly affect the statistical result)

Manual Editing

STAGE B: DATA PROCESSING

PHASE B: Completeness checks and treatment of special cases

- Checks on the number of records received
- Checks on the number of fields received
- Checks on duplicate or multiple identical records
- Treatment of duplicate or multiple identical records

PHASE C: Content checks and corrections

- Logical and accounting-mathematical checks
- Top-down scaling of errors occurred
- Manual corrections (at enterprise level) of most significant errors (in terms of value and impact)
- Automated corrections

PHASE A: Update the Business Register

- Update the BR (births and deaths of enterprises, economic activity, turnover, employment)
- Compilation of a list of Active enterprises from the updated BR

PHASE B: Compilation of SBS variables

- Compilation of SBS variables from the administrative data
- Estimation of missing variables based on a small scale survey on large enterprises



Estimation of missing variables

- The methodology applied for the estimation for the SBS variables that are not available in the administrative data is based on data derived from a small scale survey conducted on large enterprises by economic activity sector (4digit level).
- Following the compilation of SBS variables from the administrative data advanced estimation techniques are applied for the compilation of missing variables:
- Proportions of a component variable
- Predictive Models : Regression models and two part models: decision trees in combination with multiple regression models.
- ✓ Simple Ratio Adjustment Method

- STAGE D: Integration of data in the production data base for the compilation of the final results
 - In the context of the compilation of the SBS data base the files produced in the previous are integrated in the following order:
- 1. The file of active enterprises indicating the turnover and employment
- 2. The file with the rest SBS variables compiled from the administrative data
- 3. The estimated variables that are not available in the administrative data

CONCLUSIONS

- The experience from the use of administrative data derived from the tax declarations submitted by the enterprises to the Tax Authorities for the compilation of SBS was positive.
- The overall quality of the statistical output in the domain of SBS was improved significantly due to the fact that full exhaustiveness was achieved
- Acceleration of the production time with a significant reduction of human and financial resources while on the same time a significant reduction burden to the enterprises has been achieved
- Enhance further the possibility to produce modernized indicators for business statistics in order to meet the growing needs of the users



Hellenic Statistical Authority