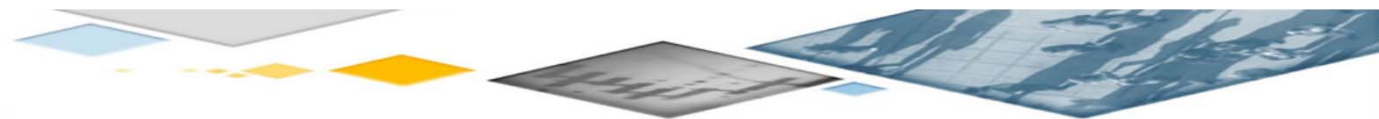




Integration of survey and administrative data in Structural Services Statistics

Laura Otero Franco, Patxi Garrido Espinosa
EUSTAT. Basque Statistics Office

European Establishment Statistics Workshop Poznań, Poland
7 September 2015



Outline

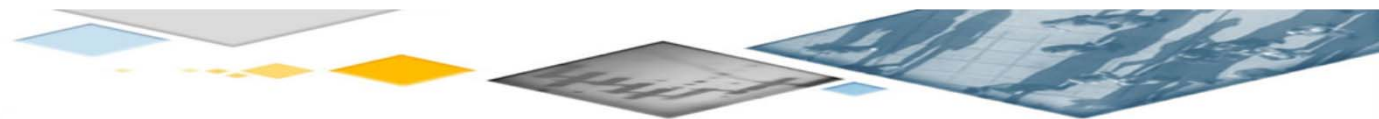
Background

New Structural Services Statistics

Data cleaning

Estimation methods

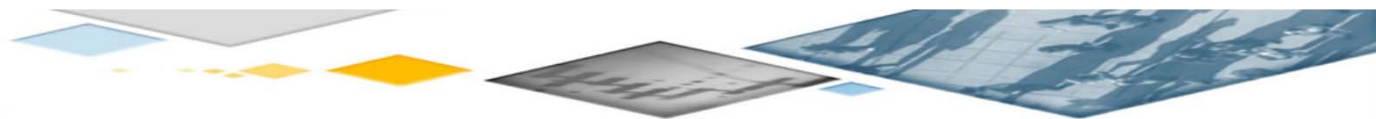
Conclusions



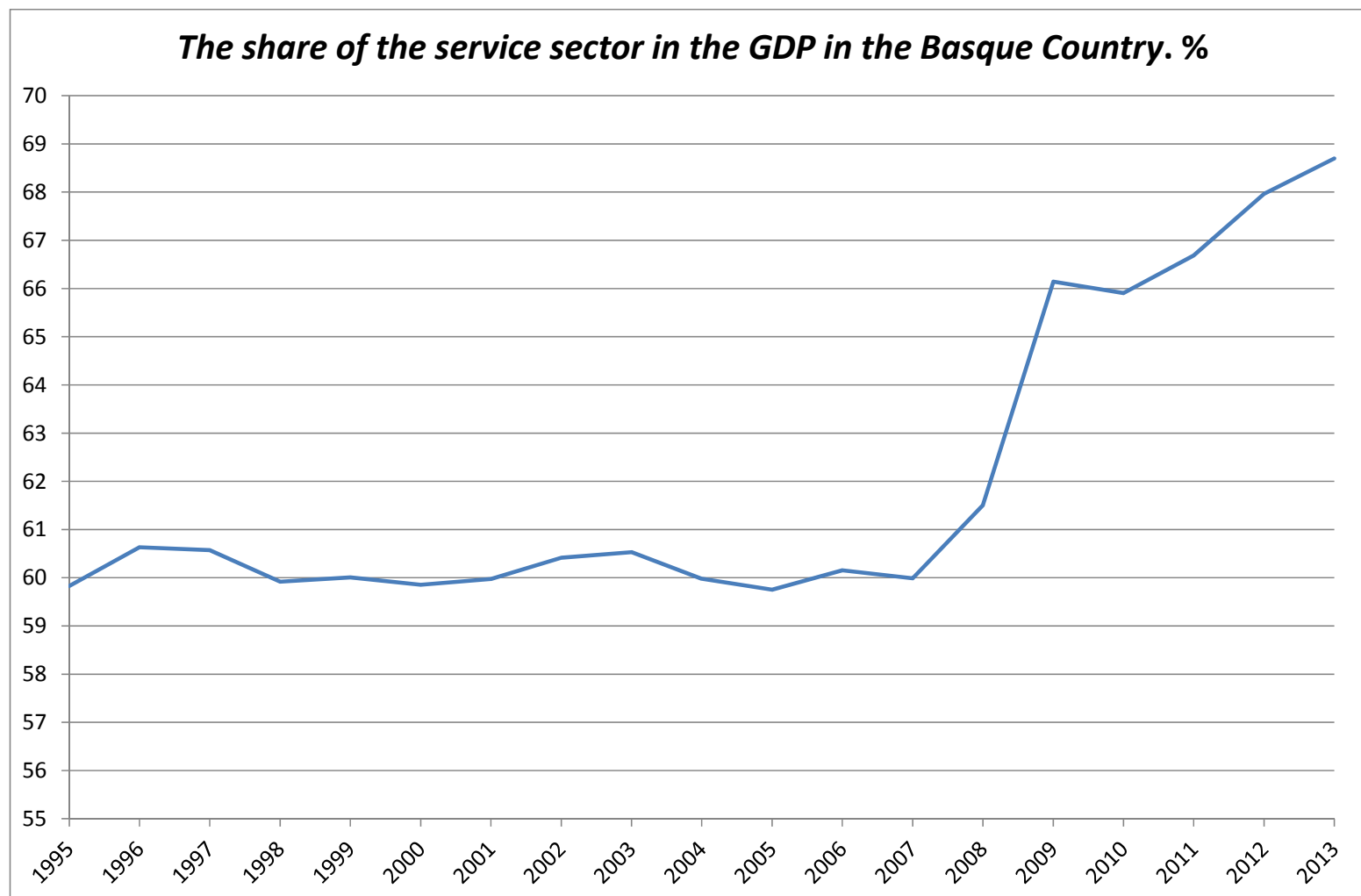
Background

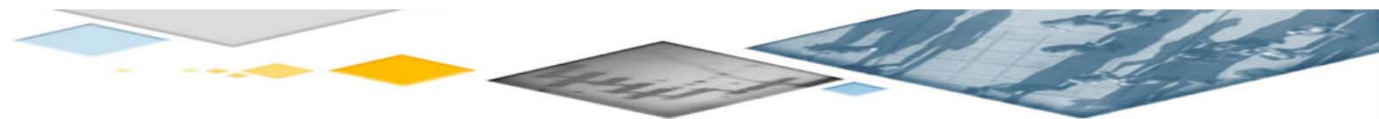
- Survey on Trade and Repairs (STR)
- Survey on Hotel and Catering Sector (SHCS)
- Survey of Other Services (SOS)

STR:	1986, 1987, 1988, 1989, 1990, 1993, 1995, 1997,	2000, 2005, 2010
SHCS:	1985,	1994, 1999, 2004, 2009
SOS:	1994, 1996, 1998, 1999, 2004, 2009	



Background





Background

Aims:

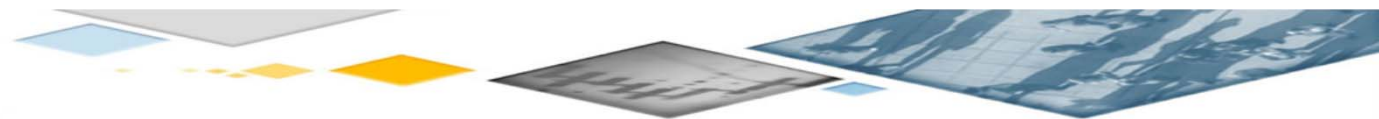
Yearly estimates

Reduce respondents burden



STRUCTURAL SERVICES STATISTICS (SSS)

SSS is carried out combining
both **admin data** and **survey data**



The new SSS: Data sources

Administrative sources

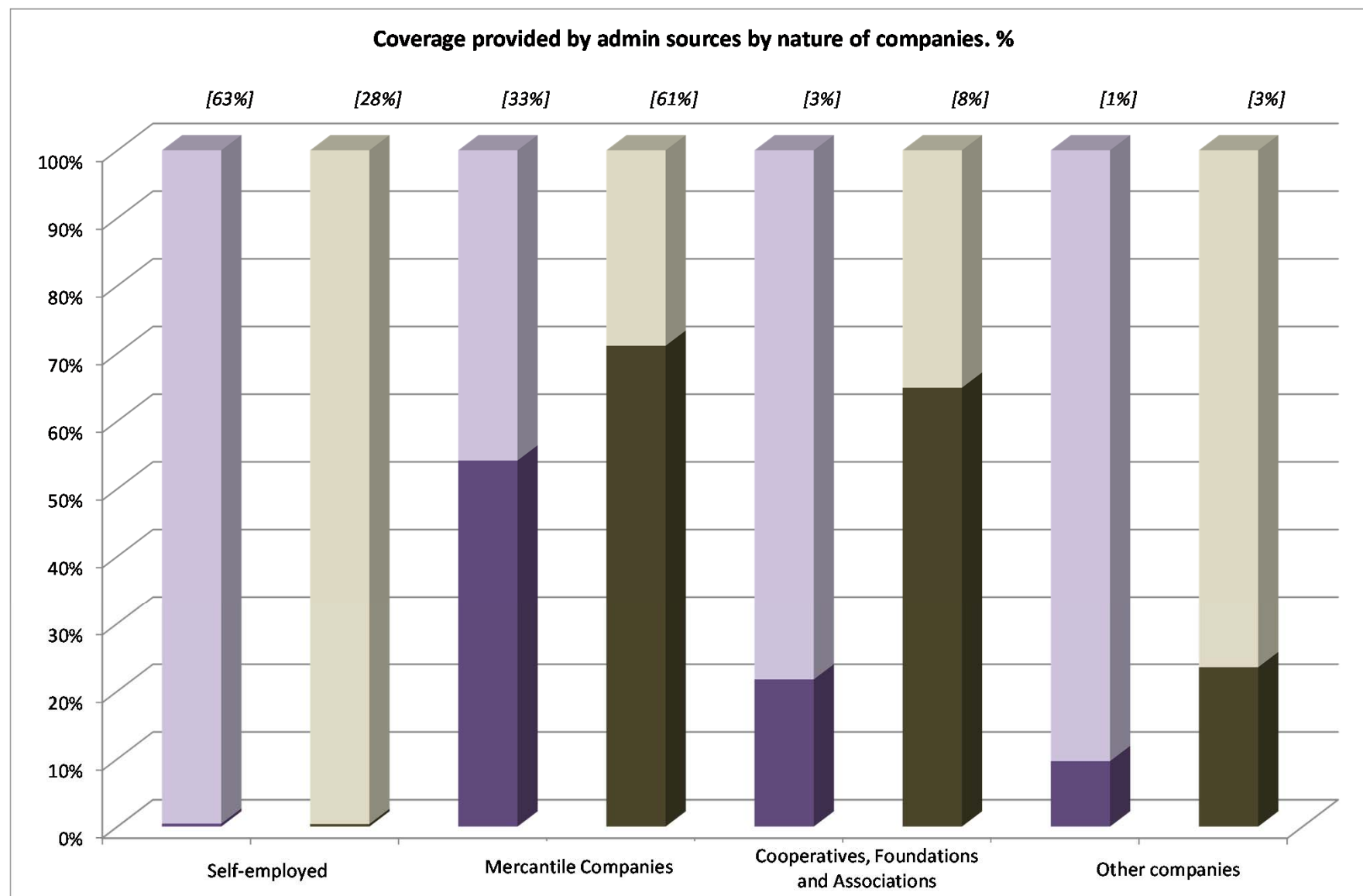
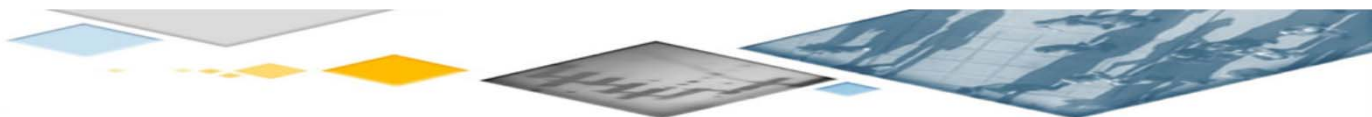
Spanish Commercial or Mercantile Register	\cong 35,000 ; company level
Basque Register of Cooperatives	60% from the Services activity
Basque Register of Foundations and Associations	Specific legal nature

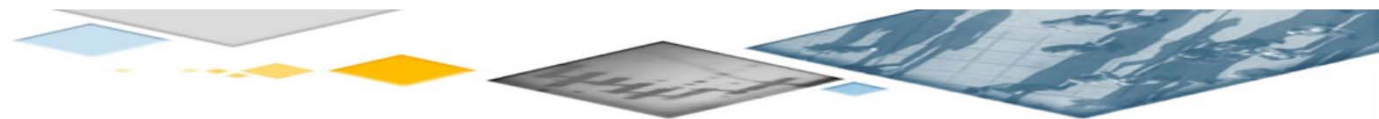


"Equity Account of Non-Financial Companies"

Statistical sources

Directory of Economic Activities (Business Register)	Population scope of all economic surveys \cong 117,500 services establishments
Services Survey	



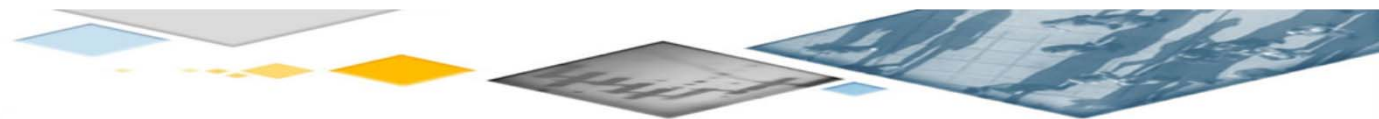


The new SSS: Sample design

Random ad-hoc sample stratified by size, legal nature and activity

ESTABLISHMENT SIZE	SAMPLING STRATEGY	SAMPLE SIZE
Self-employees	Simple Random Sampling (SRS*)	2,363 units
1-9 employees	SRS* except for mercantile companies, cooperatives, foundations and associations	1,095 units
10-49 employees	Exhaustive except for mercantile companies, cooperatives, foundations and associations	380 units
50 employees or more	Exhaustive	662 units
		4,500 units

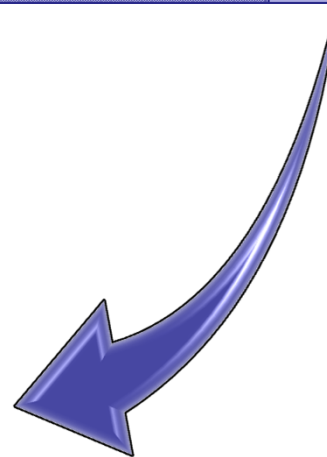
* SRS proportional to number of establishments according to observed variability for Gross Value Added (GVA) by activity



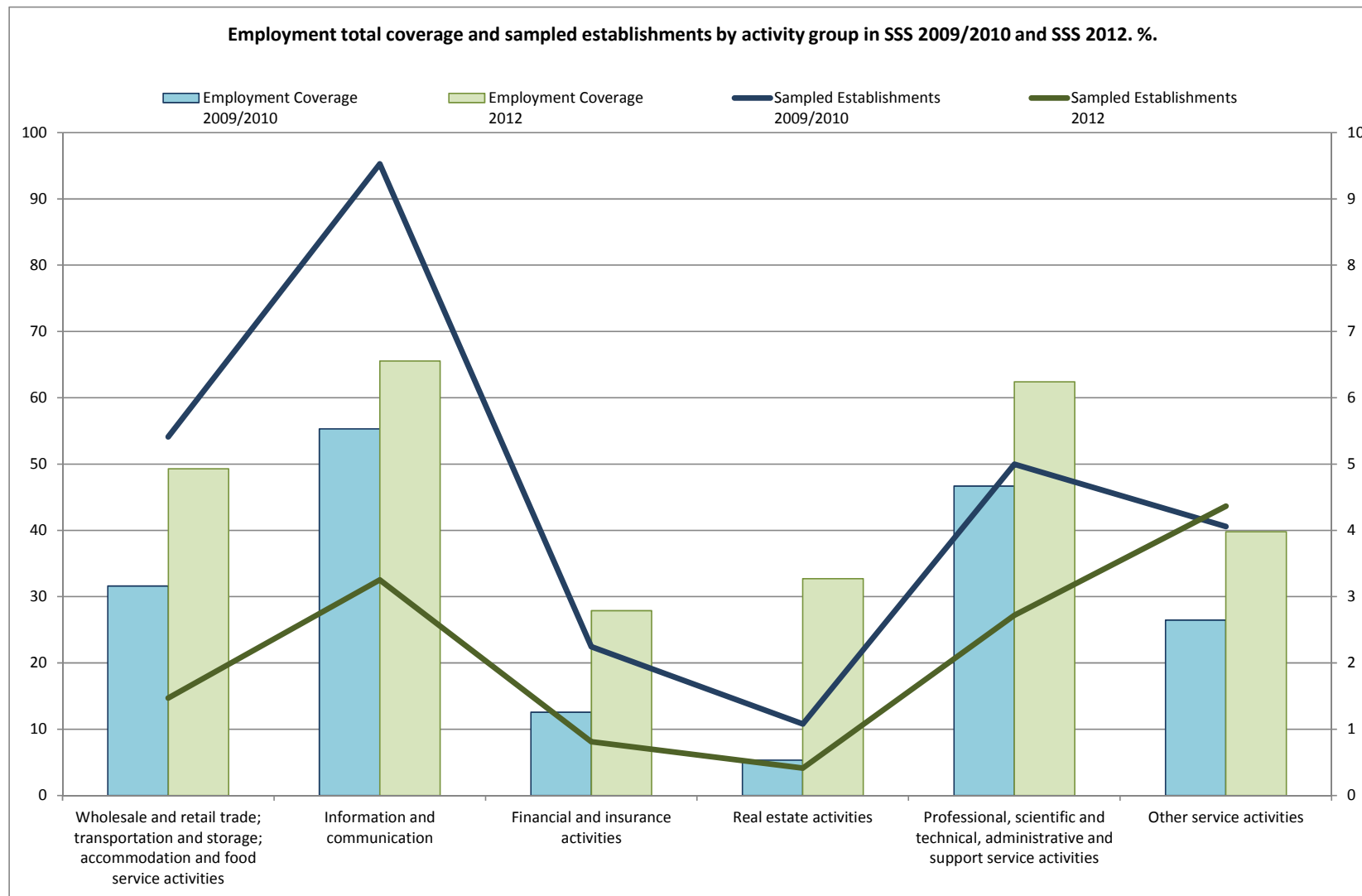
The new SSS: Sample design

Survey on Trade and Repairs (STR) 2010	3,720
Survey on Hotel and Catering Sector (SHCS) 2009	1,500
Survey of Other Services (SOS) 2009	5,500
	10,720
Structural Service Statistics (SSS) 2012	4,500
SAVING	-58%

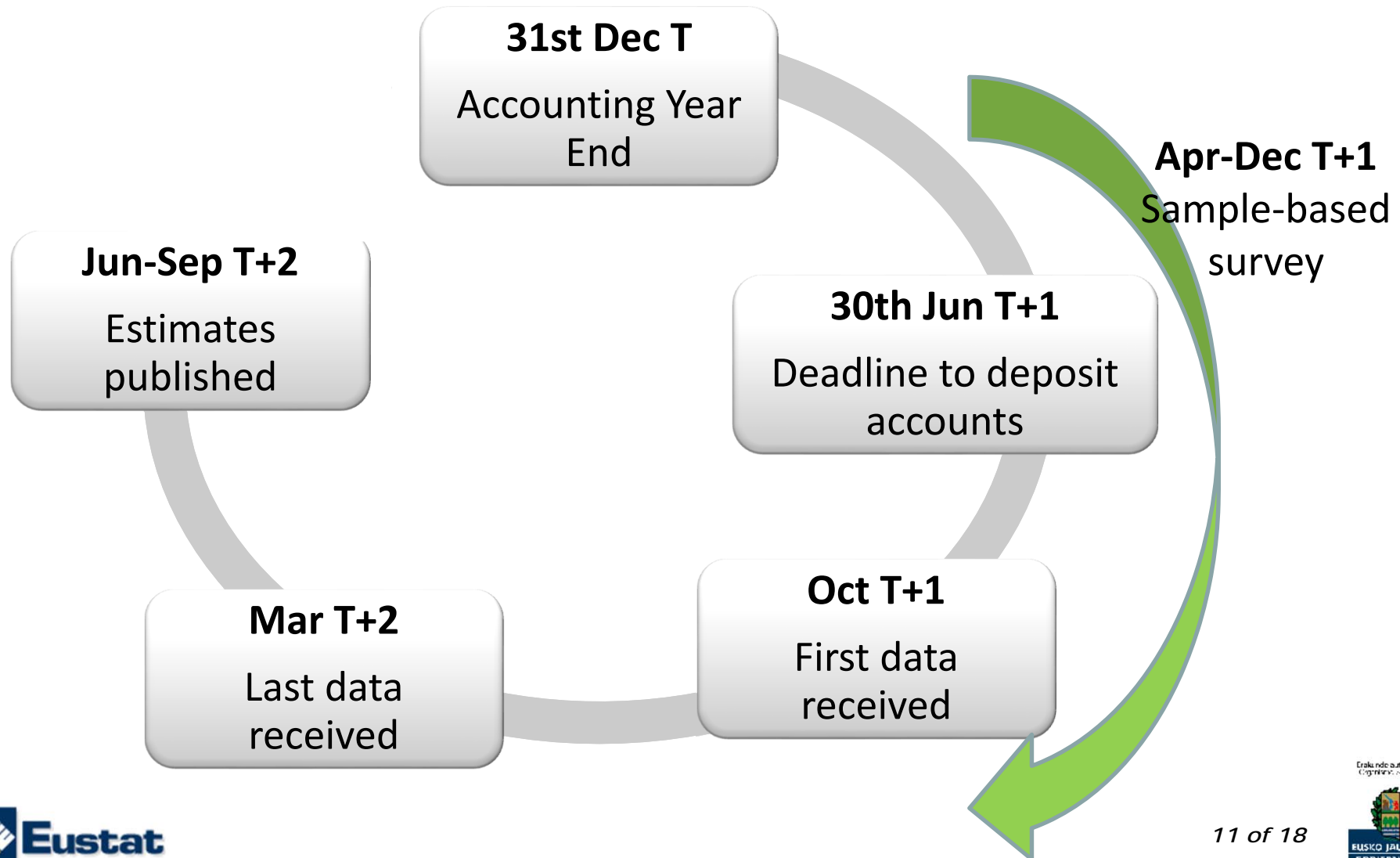
The saving does **not** imply by any means a reduction in the coverage of the population with the multi-source approach

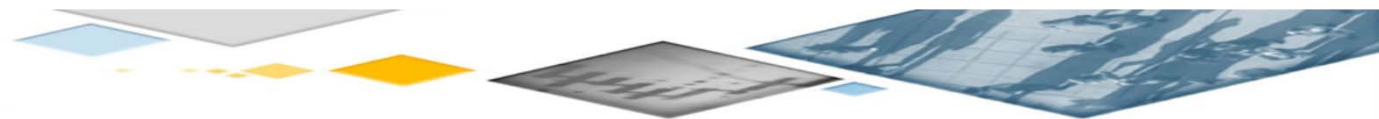


The new SSS: Effective coverage



The new SSS: The new strategy



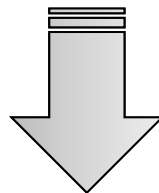


Data cleaning

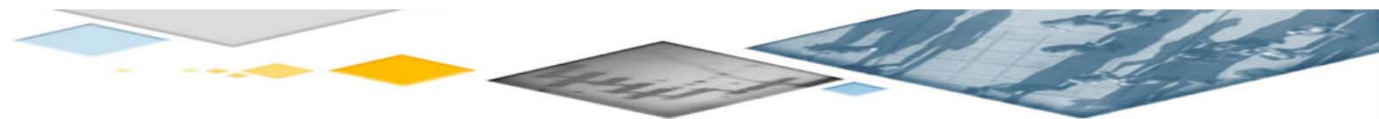
The number of information data available is considerably larger

The treatment of administrative data by the register owners does not need to follow the standards of statistical quality

The purpose of these registers is reduced to deposit the economic accounts, and there is usually no control of internal consistency and economic meaning



It is required to use automatic data cleaning techniques



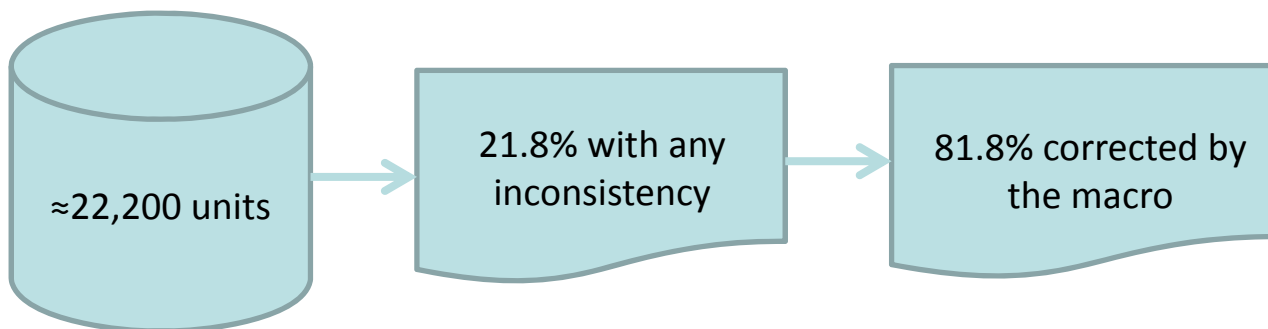
Data cleaning

1.-SAS Macro which correct economic data according with pre-established validation rules

A) Deductive imputation of breakdowns

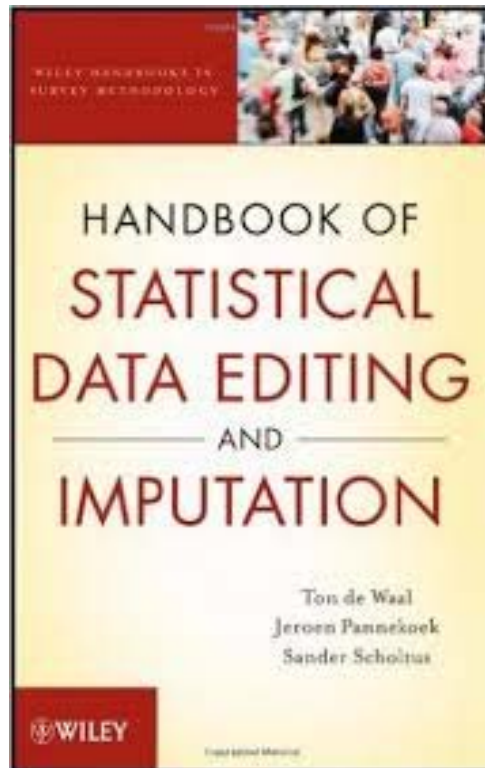
B) Correction of signs (+/-) of economics variables

C) Correction of errors due to rounding



Data cleaning

2.- SAS Macro of selective editing



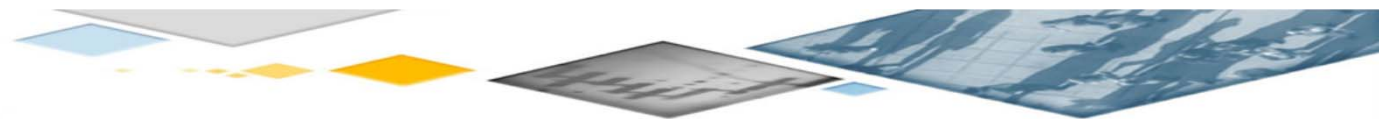
$$s_i = F_i \times R_i = \frac{w_i \tilde{x}_i}{\tilde{X}} \times \frac{|x_i - \tilde{x}_i|}{\tilde{x}_i}$$

w_i elevation weight

\tilde{x}_i real value of the variable

\tilde{x}_i expected value of the variable

\tilde{X} estimation of the total based on expected values



Data cleaning

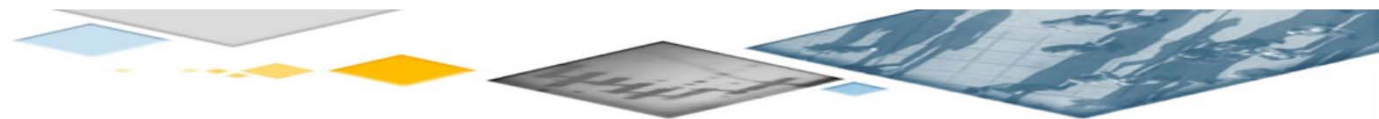
2.- SAS Macro of selective editing

Influence variables:

- Turnover
- Gross Value Added
- Wages and salaries

Risk factors (*Distances*):

- From the value of previous period
- From the value of the median
- From the expected value obtained by robust regression



Estimation methods

DIRECT VARIABLES

Composite estimator by strata using a set of weights depending on the sample fraction and using employment as auxiliary variable

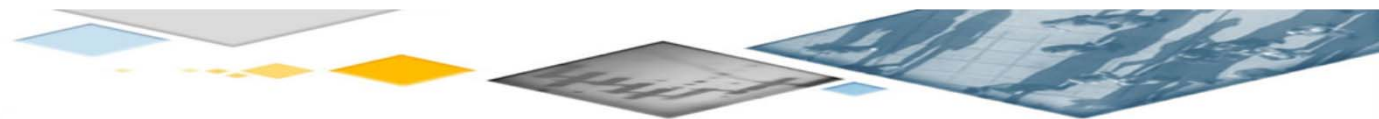


40700 OTHER OPERATING EXPENSES

40710	External services
40720	Taxes
40790	Other current management expenses

INDIRECT BREAKDOWNS

Ratio Based Imputation system using homogeneous units from a previous reference year



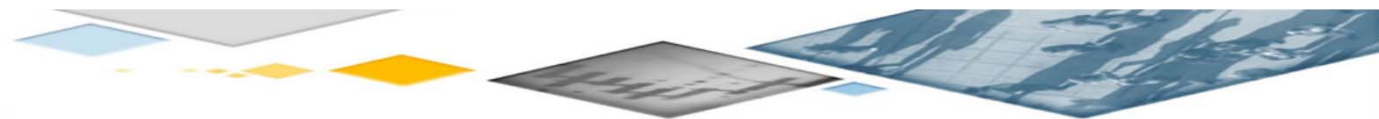
Conclusions and future work

Conclusions:

- ☐ Yearly estimates for the whole services sector together are obtained
- ☐ Respondents' burden and economic cost substantially reduced with actually a higher and better coverage

Future work:

- ☐ Minimize the overlap of admin sources and sample
- ☐ Research the indirect ratio estimators performance
- ☐ Extensive use of admin sources in all economic sectors



Muchas gracias

Thank you

Dziękuję bardzo

laura-otero@eustat.eus

patxi-garrido@eustat.eus

www.eustat.eus