



## **Quality of BCS data: Results of the Task Force on Sampling methods**



**Brussels, 14th – 15th November 2013**

# 0

---

## Background

---

## Background:

The ECFIN invited all partner institutes, taking part in the Business and Consumers surveys, to join into six different Research Task Forces, with a common goal: **Improving the quality of the European Business Cycle Indicators in the sense of accuracy of the data.** Volatility is a major indicator of that.

We decided to join into the 4th one, and we have finally been leading Task4 – Sampling methods in both groups: Business and Consumers.

## Active members of the task force:

### Consumers



Company	Country	Name of contact	Mail of contact
GfK	Germany	Klaus Hilbinger	Klaus.hilbinger@gfk.com
SL	Spain	Raquel García	rgarcia@simplelogica.com
		Eduardo Pérez	eperez@simplelogica.com



### Business



Company	Country	Name of contact	Mail of contact
Wifo	Austria	Gerhard Schwarz	gerhard.schwarz@wifo.ac.at
		Werner Hoelzl	Werner.Hoelzl@wifo.ac.at
CBI	United Kingdom	Jonathan Wood	Jonathan.Wood@cbi.org.uk
		Christopher Taylor	christopher.taylor@cbi.org.uk
CBM	Montenegro	Veljko Rudanovic	veljko.rudanovic@cb-cg.org
SL	Spain	Raquel García	rgarcia@simplelogica.com
		Eduardo Pérez	eperez@simplelogica.com



Thank you for your valuable contribution.

# 1

---

## Goals and Hypothesis

---

## Goals

**“The final goal of the task force will be to identify best practices and “criteria for success” in carrying out business and consumer surveys”.**

And this means, in our Task Force:

Finding out whether there is any statistically significant relationship between each sampling method implemented and the quality of data, in terms of volatility and bias.

## Research hypothesis:

There are differences between the averages of volatility for each type of sampling method.

# 2

---

## Data Collection

---

## Building up the needed data set:

1<sup>st</sup>. The Metadata checked by partners" excel document (sent by the ECFIN)

	A	B	C	E	F
1	<b><u>CONSUMERS SURVEY</u></b>				
2					
3	<b>Country</b>	<b>Sampling frame</b>	<b>Size of the actual frame list</b>	<b>Population (Universe)</b>	<b>Sampling method</b>
9	CZ (GfK)	Database of households phone numbers	1.534.728	10.505.445	Stratified systematic random sampling
10	DK (DST)	CPR (Central Person Register)	4.102.093	5.560.628	Random sample by age.
11	EE (KI)	Estonian Register of residential telephone lines	300.000	1.100.000	Random sampling + Quotas for sample (quota for region, sex, age)
	FI (STAT)	Total population data base	5.400.000	4.400.000	Systematic random sampling: >>sorting of the sampling frame based on domicile code (thus yielding implicit stratification according to geographical population density).

## Building up the needed data set:

	A	B	C	E	F
1	<b><u>RETAIL TRADE SURVEY</u></b>				
2					
3	Country	Sampling frame	Size of the actual frame list	Population (Universe)	Sampling method
9	CZ (CZSO)	Register of Economic Subjects	2.100	295.700	stratified panel - criteria: >>turnover >>branch
10	DK (DST)	Establishment related employment Statistics (ERE)	4.279	28.508	stratified random sampling stratification by: >>activity >>size groups (in terms of employees) >>inclusion of all firms >5 employees
11	EE (KI)	Estonian Business Register	730	6.700	panel: >>quotas >>random sampling
12	FI (EK)	Register covering 75 % of the total turnover in retail trade in Finland	12.000		random sampling: >>all large firms included

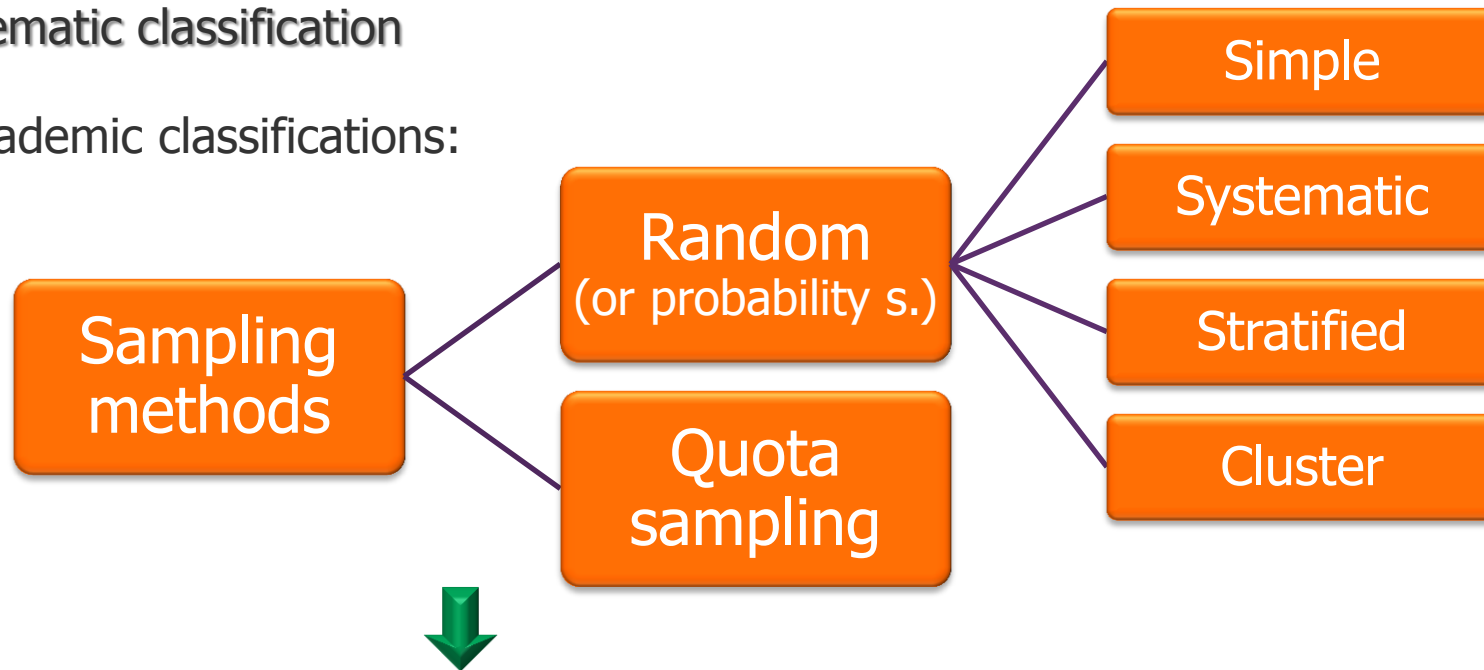


## Building up the needed data set:

It is necessary to be sure that every partner institute is answering using a:

- common and
- systematic classification

Usual academic classifications:



Our own classification of sampling methods was built up.

## Building up the needed data set:

Our own classification of sampling methods was built up in a questionnaire:

### 1. Universe inclusiveness

- **EXHAUSTIVE sampling:**  
*Could anyone in the Universe be selected?*
- **LIMITED or FOCUSED sampling:**  
*Do you focus in any specific target? (f.example: large companies or citizens in middled-large cities?)*
- **PANEL:**  
*Do you conduct your survey using a panel?*

### 2. Sampling units selection

- **SIMPLE RANDOM Sampling:**  
*Do you do it (having access to 100% of universe, in order to select randomly 1 in x)?*
- **STRATIFIED Sampling:**  
*Do you distribute the Universe by any variable, such as: NACE, turnover, Region or town-size for example?*
- **DIRECTLY PROPORTIONAL stratified sampling:**  
*Is it directly proportional to the Universe distribution?*

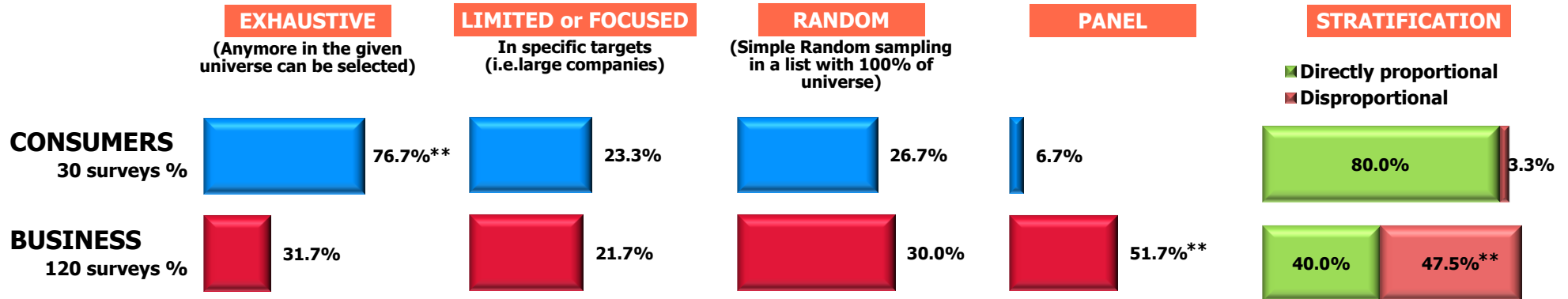
## Building up the needed data set:

Our own classification of sampling methods was built up in a questionnaire:

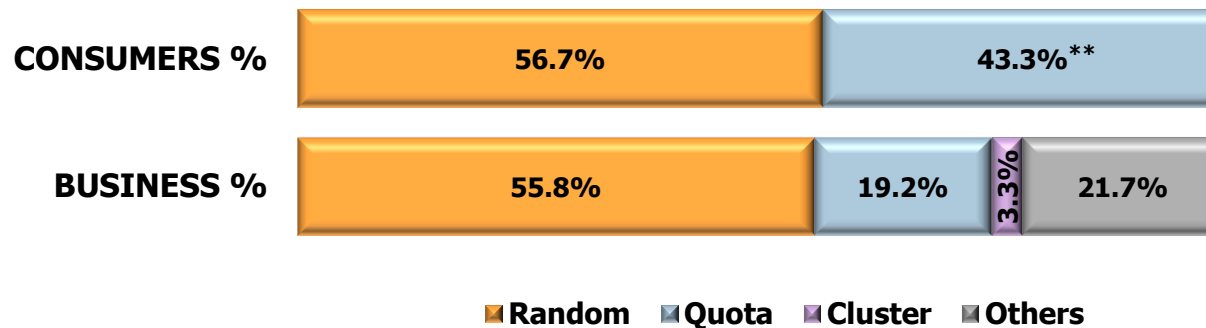
### 3. Final respondent selection

- ON RANDOM
- QUOTAS
- CLUSTER  
*(All individuals within the cluster are interviewed)*

## Sampling methods implemented, by target (In % of surveys\*)



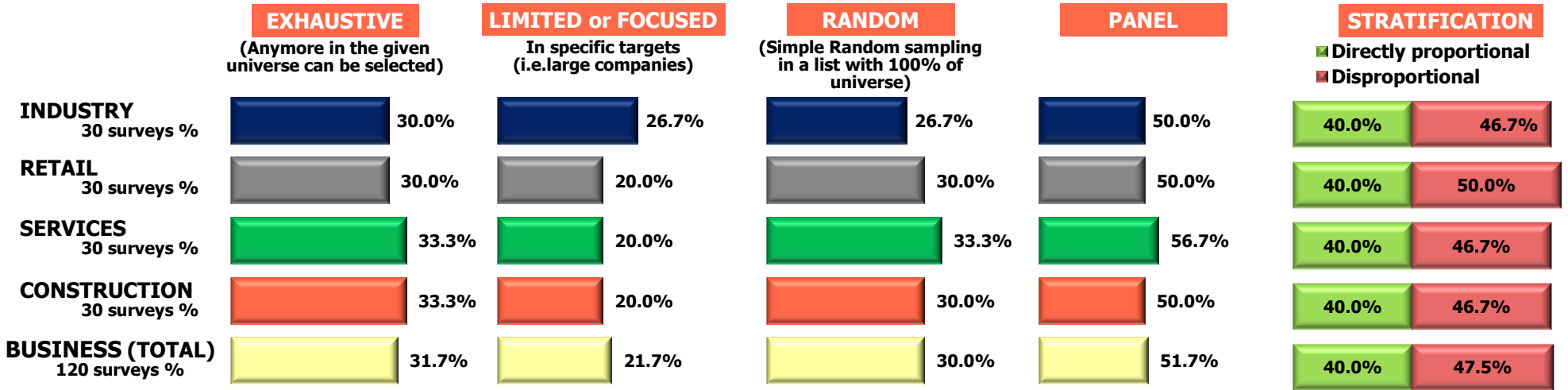
## Final respondent selection



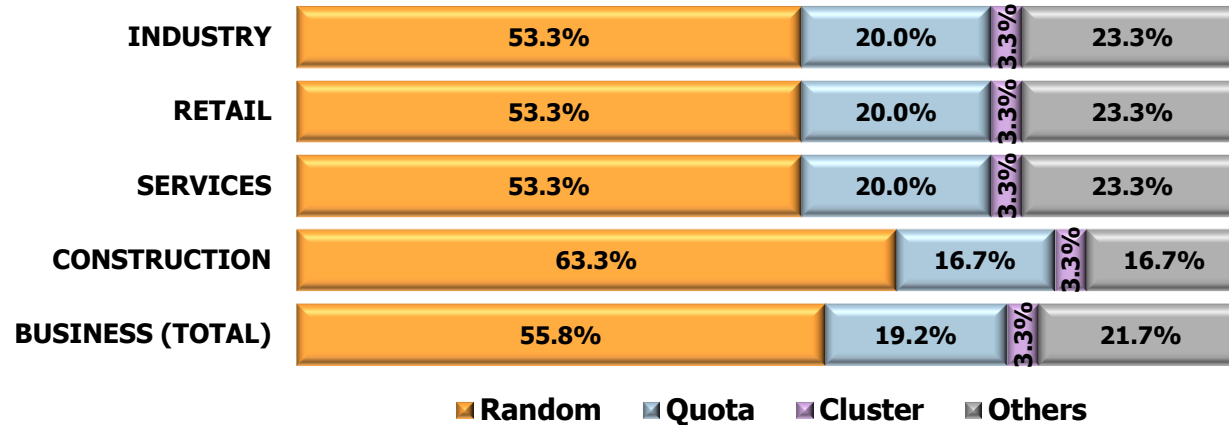
\* Business surveys: 30 countries x 4 surveys types = 120

\* Consumer surveys: 1 per country

## Sampling methods implemented in Business, by survey (In % of surveys\*)



## Final respondent selection



\*Survey: 30 countries x 4 surveys types = 120 Business surveys.

# 3

---

## Analysis

---

## Analysis:

Three steps of analysis were done:

**A.**

• Correlation

**B.**

• Regression

**C.**

• Chi<sup>2</sup> of Mean scores

## Analysis: Step A - Correlations

### BUSINESS

		EXHAUSTIVE_ SAMPLING	LIMITED_OR_ FOCUSED	SIMPLE_RANDOM_ SAMPLING	PANEL	STRATIFIED	STRATIFIED_ PROPORTIO	SELECTION_ FINAL_RESP
VOLATILITY (MCD)	Pearson Correlation	-,024	-,143	-,071	-,046	,046	-,055	,058
	Sig. (bilateral)	,812	,153	,483	,646	,651	,585	,565
	N	101	101	101	101	101	101	101

### CONSUMERS

		EXHAUSTIVE_ SAMPLING	LIMITED_OR_ FOCUSED	SIMPLE_RANDOM_ SAMPLING	PANEL	STRATIFIED	STRATIFIED_ PROPORTIO	SELECTION_ FINAL_RESP
VOLATILITY (MCD)	Pearson Correlation	,124	-,124	-,004	-,059	,236	,112	-,161
	Sig. (bilateral)	,537	,537	,982	,768	,237	,577	,422
	N	27	27	27	27	27	27	27



## Analysis: Step B - Regressions

**Model summary (BUSINESS)**

Model	R	R square	R square corrected	Standard error of the stimation
	,143	,031	,021	,974

**Model summary (CONSUMERS)**

Model	R	R square	R square corrected	Standard error of the stimation
	,236	,055	,018	,712

## Analysis: Step C - Chi<sup>2</sup> of Mean scores

Exhaustive in business/Consumers				
	Exhaustive_Business	Exhaustive_Consumers	NonExhaustive_Business	NonExhaustive_Consumers
	Mean	Mean	Mean	Mean
VOLATILITY	3,00a	3,20a	3,05a	3,00a

Limited and business/Consumers				
	Limited_Business	Limited_Consumers	NonLimited_Business	NonLimited_Consumers
	Mean	Mean	Mean	Mean
VOLATILITY	2,78a	3,00a	3,12a	3,20a

Simple random in business/Consumers				
	SimpleRandom_Business	SimpleRandom_Consumers	NonSimpleRandom_Business	NonSimpleRandom_Consumers
	Mean	Mean	Mean	Mean
VOLATILITY	2,93a	3,14a	3,08a	3,15a

Panel in business/Consumers				
	Panel_Business	Panel_Consumers	NonPanel_Business	NonPanel_Consumers
	Mean	Mean	Mean	Mean
VOLATILITY	3,00a	3,00a	3,09a	3,16a

Stratified in business/Consumers				
	Estratified_Business	Estratified_Consumers	NonEstratified_Business	NonEstratified_Consumers
	Mean	Mean	Mean	Mean
VOLATILITY	3,06a	3,23a	2,93a	2,80a

Estratified proportional in business/Consumers				
	EstratProp_Business	EstratProp_Consumers	NonEstratProp_Business	NonEstratProp_Consumers
	Mean	Mean	Mean	Mean
VOLATILITY	2,98a	3,19a	3,08a	3,00a

## Analysis: 2<sup>nd</sup> ROUND

### Features:

- ✓ Include in the analysis other indicators of volatility (MCD=1 or MCD=2) and the tracking performance or bias of the survey results (as measured by the correlations between confidence indicators and statistical reference series)
- ✓ Construct variables of sampling methods avoiding overlapping

## Analysis: 2<sup>nd</sup> ROUND

### 2<sup>nd</sup> feature: Construct variables of sampling methods avoiding overlapping

#### For business surveys:

- **Panel\_business:** *Is business survey using a panel or something different?*
  - 1 = business survey is using a panel
  - 0 = business survey is using something else (random sampling, quota sampling, etc.)
- **Limited\_business:** *Is the business survey limited / focused?*
  - 1 = business survey is limited / focused (e.g. focus on large companies)
  - 0 = business survey NOT limited or focused.
- **Random\_strat\_business:** *Does the business survey, which is based on random sampling, use stratification or not?*
  - 1 = business survey is using random sampling **and** using stratification
  - 0 = business survey is using random sampling, but is NOT using stratification.

Note: only 28 observations <-> 101 at the other two new variables

#### For consumer surveys:

- **Panel\_consumer:** *Is consumer survey using a panel or something different?*
  - 1 = consumer survey is using a panel
  - 0 = consumer survey is using something else (random sampling, quota sampling, etc.)
- **Limited\_consumer:** *Is the consumer survey limited / focused?*
  - 1 = consumer survey is limited / focused (e.g. focus on middle and large cities)
  - 0 = consumer survey NOT limited or focused.
- **Random\_strat\_consumer:** *Does the consumer survey, which is based on random sampling, use stratification or not?*
  - 1 = consumer survey is using random sampling **and** using stratification
  - 0 = consumer survey is using random sampling, but is NOT using stratification.

Note: only 7 observations <-> 27 at the other two new variables

## 2<sup>nd</sup> Analysis: Step A - Correlations

Correlations:		BUSINESS			CONSUMERS		
		panel_business	limited_business	random_strat_business	panel_consumer	limited_consumer	random_strat_consumer
VOLATILITY (MCD)	Pearson correlation	-,046	-,143	-,038	-,059	-,124	,545
	Sig. (bilateral)	,646	,153	,849	,768	,537	,206
	N	101	101	28	27	27	7
MCD=1	Pearson correlation	-,120	-,217	,128	,064	,038	-,178
	Sig. (bilateral)	,231	,029	,516	,752	,850	,703
	N	101	101	28	27	27	7
MCD=2	Pearson correlation	-,081	-,125	-,104	-,089	-,105	,666
	Sig. (bilateral)	,419	,211	,599	,659	,601	,103
	N	101	101	28	27	27	7
CORRELATION (Survey indicator performance)	Pearson correlation	-,136	,063	-,183	-,060	,160	-,182
	Sig. (bilateral)	,201	,552	,361	,767	,426	,696
	N	90	90	27	27	27	7

## 2<sup>nd</sup> Analysis: Step B – Regressions (Volatility)

### BUSINESS (MCD=1)

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
panel_business limited_business	,239	,057	,038	,929

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
random_strat_business	,128	,016	-,021	1,028

### BUSINESS (MCD=2)

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
panel_business limited_business	,143	,021	,001	,473

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
random_strat_business	,104	,011	-,027	,440

### CONSUMERS (MCD=1)

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
panel_consumer limited_consumer	,064	,004	-,079	,634

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
random_strat_consumer	,178	,032	-,162	,408

### CONSUMERS (MCD=2)

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
panel_consumer limited_consumer	,114	,013	-,069	,374

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
random_strat_consumer	,666	,443	,332	,351

Model	Sum of squares	df	Root-mean-square	F	Sig.
Regression	,489	1	,489	3,977	,103
Residual	,615	5	,123		
Total	1,104	6			

## 2<sup>nd</sup> Analysis: Step B – Regressions (Bias)

### BUSINESS (Correlation: Survey indicator performance)

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
panel_business limited_business	,162	,026	,004	21,169

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
random_strat_business	,183	,033	-,005	24,768

### CONSUMERS (Correlation: Survey indicator performance)

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
panel_consumer limited_consumer	,223	,050	-,030	23,928

Model summary

Model	R	R square	R square corrected	Standard error of the stimation
random_strat_consumer	,182	,033	-,160	21,084

## 2<sup>nd</sup> Analysis: Step C - Chi<sup>2</sup> of Mean scores

### BUSINESS

	panel_business	
	No	Yes
	Mean	Mean
MCD=1	2,47a	2,24a
MCD=2	1,32a	1,24a
SURVEY INDICATOR PERFORMANCE	68,68a	62,85a

	limited_business	
	No	Yes
	Mean	Mean
MCD=1	2,45a	1,96b
MCD=2	1,30a	1,16a
SURVEY INDICATOR PERFORMANCE	64,53a	67,75a

	random_strat_business	
	No	Yes
	Mean	Mean
MCD=1	2,09a	2,43a
MCD=2	1,33a	1,21a
SURVEY INDICATOR PERFORMANCE	71,00a	58,52a

### CONSUMERS

	panel_consumer	
	No	Yes
	Mean	Mean
MCD=1	2,21a	2,36a
MCD=2	1,28a	1,16a
SURVEY INDICATOR PERFORMANCE	58,28a	53,00a

	limited_consumer	
	No	Yes
	Mean	Mean
MCD=1	2,21a	2,26a
MCD=2	1,29a	1,21a
SURVEY INDICATOR PERFORMANCE	55,70a	64,14a

	random_strat_consumer	
	No	Yes
	Mean	Mean
MCD=1	2,41a	2,29a
MCD=2	1,00a	1,54a
SURVEY INDICATOR PERFORMANCE	57,00a	50,33a





# 4

---

## Conclusions

---

## Conclusions

### **Impact of sampling methods on volatility and bias**

- There is no clear evidence of impact of any of the different sampling methods on volatility (MCD). Nor in the consumer nor in the business surveys either.
- The only result we are able to obtain is that limited sampling is associated with a lower volatility score in the business surveys for  $MCD=1$ . That is, the surveys which are focused on specific sorts of companies seem to have lower volatility scores.  
However, even this result is very weak.
- No evidence of impact has been found at all between any sampling method and bias (the correlation coefficient between survey confidence indicators and its reference series).
- The sampling methods being mutually exclusive or being overlapped, do not change these results.



---

**Thank you!**

---



## Quality of BCS data: Results of the Task Force on Sampling methods



Simple Lógica: Raquel Garcia: [rgarcia@simplelogica.com](mailto:rgarcia@simplelogica.com)  
Eduardo Perez: [eperez@simplelogica.com](mailto:eperez@simplelogica.com)