FROM AVERAGE TO VARIABILITY AND DISTRIBUTION-BASED INDICATORS IN BUSINESS STATISTICS

WHAT POLICY MAKERS CAN LEARN FROM OFFICIAL STATISTICS ABOUT STRUCTURAL TRENDS IN THE BUSINESS ECONOMY?

UN Committee of Experts on Business and Trade Statistics

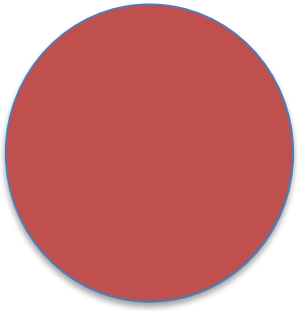
11-13 June 2019 - New York

Alessandro Faramondi and Stefano Menghinello
Directorate for Business Statistics
Italian National Institute for Statistics (ISTAT)

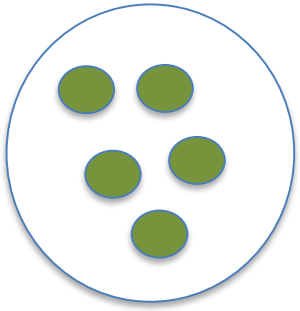
Goal of the presentation

- Let's start with a **question/little game**: can you identify "winners" and "loosers" of economic growth in your country by naming few industries, a specific company size such as SME versus large ones, location in specific regions?
- Why? **Standard economic classifications** (industry, enterprise size class and location) **are less and less capable to capture the complexity of businesses**. There is a need to detect the new drivers of growth (global and digital enterprises) and to mitigate uneven growth
- How to fix it? By leveraging methodological innovation in business statistics, it is possible to introduce a **new class of indicators that can provide a broader and more consistent picture of nowadays structure and evolution of businesses**

The new approach to SBS: a process and product innovation



The business register



Model A
Random based direct reporting approach

Model B
Register based administrative data intensive approach

THE SHIFT FROM MODEL A TO B AS SBS PRODUCTION FRAMEWORK HAS GENERATED TWO IMPORTANT RESULTS:

- MORE ACCURATE AND DETAILED SBS DATA (NO SAMPLING ERRORS AND CONTRAINTS)
- THE OPPORTUNITY TO USE NON STANDARD CLASSIFICATIONS AND TO EXPLORE FIRM LEVEL VARIABILITY

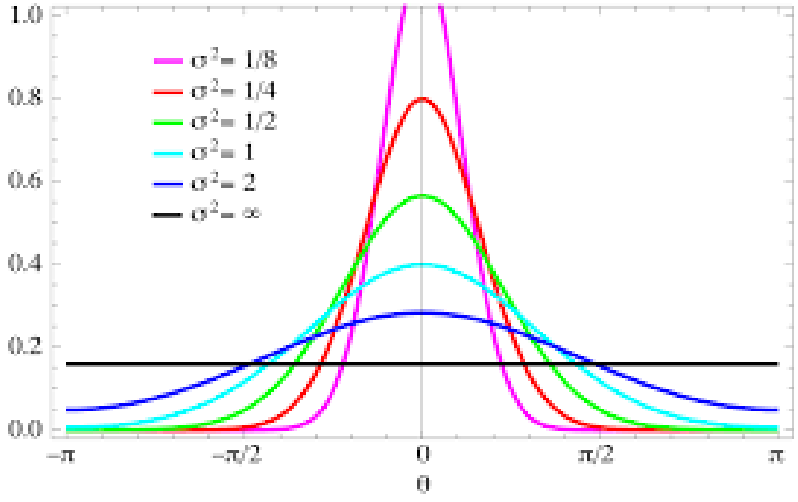
The new approach to SBS: a process and product innovation

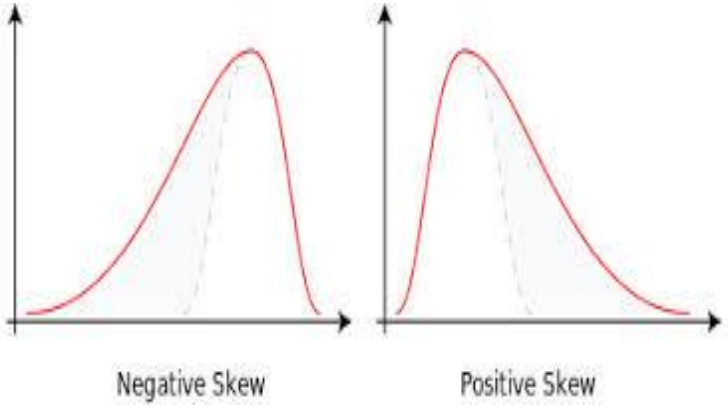
- THE **TRADITIONAL FRAMEWORK** FOR THE PRODUCTION OF SBS DATA ESSENTIALLY GENERATES TOTAL AND MEAN VARIABLES PLUS ALLOWS FOR MODEL-BASED ANALYSIS OF FIRM LEVEL HETEROGENEITY. **ONLY THE DISTRIBUTION OF FIRMS IN THE SAMPLE IS KNOWN.**

- THE **NEW SBS FRAMEWORK** ALLOWS TO FULLY EXPLOIT THE DATA ALONG DIFFERENT STANDARD AND NOT STANDARD DIMENSIONS AND TO EXPLORE FIRM LEVEL HETEROGENEITY IN A FULLY CONSISTENT WAY WITH RESPECT TO OFFICIAL FIGURES: **THE FINITE DISTRIBUTION OF THE WHOLE TARGET POPULATION IS KNOWN. GENUINE VERSUS "ARTIFICIAL" FIRM LEVEL VARIABILITY OF KEY VARIABLES MUST BE CAREFULLY MANAGED.**

Why variability and distribution-based indicators are important

Let's think about the firm level productivity (log) distributions in the same industry and size class but across different countries



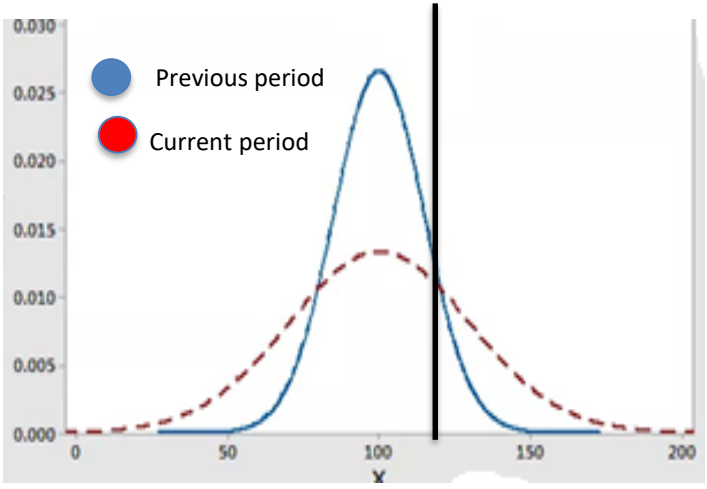


Since all distributions share the same «mean» value for productivity, **traditional SBS figures are unable to depict differences in terms of firm heterogeneity**. Some countries (lower variance) have more similar firms, others (higher variance) have a business community more polarised in extreme positions (strong winners but also firms with severe problems)

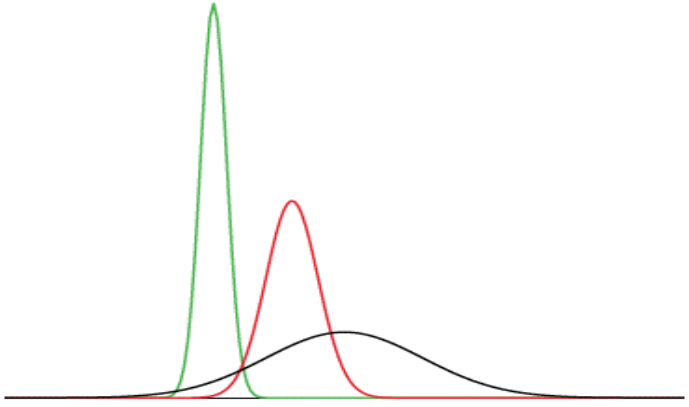
Also **the assumption of a fully symmetrical distribution of firms productivity is quite unrealistic**. Best performing companies are rare, while firms with severe problems are relatively more common

Why variability and distribution-based indicators are important

Similar considerations apply if we consider the evolution over time of a firm level productivity (log) distribution for a given country, industry and size class



Productivity stagnation (no change in mean values over time) hides a **strong increase in variability:** best performing firms (global and digital ones) are more competitive but their impact on firms mean productivity is nullified by increasing difficulties for lagging ones.



Productivity evolution over time (change of both mean and variability over time). **Standard business statistics only captures change in mean** (increase of productivity) while information on increased variability / dispersion is missing.

Why variability and distribution-based indicators are important

So how to provide a broader picture on the evolution of industries and firms ?

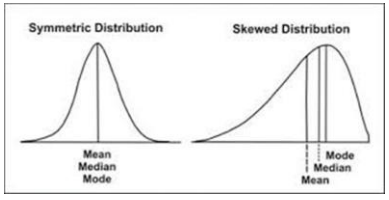
A new set of indicators calculated from micro-level data shell be added as a standard output to the traditional set of SBS variables and indicators for each domain (industry, size class, region) and their combination

Type one: measure of the variability of key indicators (Labour productivity level)

For instance: **coefficient of variation (CV)** is defined as the ratio of the standard deviation to the mean

Type two: different measures of "mean" and "asymmetry"

For instance: **mean, median mode and their "distances"** are informative on the structure and evolution of the "shape" of a distribution



Type three: measure of position for best performing, above average, below average and lagging behind firms

For instance: **computation of quartiles for a frequency distribution** of productivity as position indicator for different profiles of company performance 1°, 2°, 3° and 4°

standard classifications and firm heterogeneity

If a **classification** of businesses by economic activity or by size class or by enterprise location explain a great share of firm level variability, for instance in productivity level, the classification is **business relevant**, indeed it helps to explain in a simple way the complexity of the business structure and economic performance of a country.

The **explained variability** is usually called **between class variability** in contrast to **within-class or unexplained variability**. A good classification shell maximise explained variability (or minimize unexplained variability).

The **current revision of ISIC-NACE** is finalized to improve the power of those classification not only to identify more accurately new and old economic activities but also to better explain the overall variability of an economic system in terms of value added, productivity and employment levels and growth rates

standard classifications and firm heterogeneity

Let's us consider up to what extent current economic classification schemes can explain the heterogeneity of firm- level productivity in manufacturing in Italy based upon Census-like micro-data used to produce SBS official figures

A multi-ways ANOVA model is used to decompose with and between-class variability of the (log) **labour productivity level of Italian manufacturing firms in 2016**

Variability sources	Share in % of total variability	
explained by the model, of which	16,5	21,0
economic activity	7,6	12,6
size class	5,8	5,7
unexplained	83,5	79,0
total	100,0	100,0
number of firms	370.000	62.000
size	all	>10 people

Standard classifications explain up to 21% of firm level variability in productivity of which 12,6% is explained by industry and 5,7% by company size

standard classifications and firm heterogeneity

So what's happens if **we add new classification schemes** based upon firm internationalization profiles?

- **Exporting companies**
- **Affiliates of foreign-controlled multinational enterprises**

Variability sources	Share in % of total variability	
explained by the model, of which	24,0	30,0
economic activity	5,0	9,0
size class	1,0	1,1
exporting	2,2	1,6
importing	1,8	2,8
affiliate of a foreign MNE	0,0	0,2
unexplained	76,0	70,0
total	100,0	100,0
number of firms	370.000	62.000
size	all	>10 people

New classification schemes explain by themeselves 18% of total variability. However, because of **spurious correlation** with other variables, they are considered as additional variables to the previous model.

The **explained variability moves up from 16,5% to 24% and from 21% to 30%** showing that they contribute to understand productivity heterogeneity

The evolution over time of industry and firm level productivity

Description NACE	mean	cv	skewness
Manufacture of coke and refined petroleum products	108,6	1,3	8,0
Manufacture of pharmaceutical products	134,5	1,1	4,0
Manufacture of wearing apparel	31,8	1,0	6,3
Manufacture of chemicals and chemical products	91,9	1,0	8,5
Manufacture of electrical equipment	56,9	0,9	20,2
Manufacture of food products	52,7	0,9	6,5
Manufacture of other transport equipment	52,8	0,9	3,2
Manufacture of other non-metallic mineral products	53,7	0,8	10,3
Manufacture of leather and related products	40,4	0,8	3,3
Total	**55,1**	**0,8**	**8,2**
Manufacture of beverages	87,3	0,8	2,8
Other manufacturing	48,4	0,7	4,0
Manufacture of computer, electronic and optical product	62,2	0,7	2,5
Manufacture of basic metals	65,7	0,7	-3,9
Manufacture of furniture	44,5	0,7	1,9
Manufacture of tobacco products	72,9	0,7	0,1
Manufacture of textiles	50,4	0,6	3,2
Manufacture of wood and of products	43,9	0,6	9,1
Manufacture of motor vehicles, trailers and semi-trailers	60,4	0,6	2,3
Manufacture of rubber and plastic products	61,7	0,6	5,9
Manufacture of paper and paper products	60,9	0,6	2,8
Manufacture of machinery and equipment n.e.c.	66,2	0,5	2,5
Printing and reproduction of recorded media	49,8	0,5	2,7
Repair and installation of machinery and equipment	49,6	0,5	3,3
Manufacture of fabricated metal products	54,5	0,5	2,4

Position and variability indicators of firm-level labour productivity in manufacturing for companies in Italy with at least 10 people - 2016

2 digits NACE Industries (ranked by CV) exibit a quite different degree of within industry labour productivity variability

Asymmetry in the distribution is right skewed as expected (winners are few…)

The evolution over time of industry and firm level productivity

Firm level labour productivity of Italian manufacturing companies with at least 10 person employed over the period 2012-2016

Description NACE	average	dispersion	worse (25°)	median	best (75°)
Manufacture of textiles	26,0	-22,2	27,8	19,1	18,2
Manufacture of electrical equipment	16,2	28,4	14,2	13,0	13,8
Manufacture of food products	15,0	2,4	7,7	10,1	13,7

- In the **textile industry**, labour productivity level increased faster and dipersion strongly mitigated since «worse» performing firms not only catched up but also reduced their productivity negative gap as compared to both «median» and best performing ones
- In **electrical equipment industry** both labour productivity and disperision increased, «worse» firms catched up but they didn't reduce the negative gap
- In the **manufacture of food products**, labour productivity increased slower with a limited increase in dispersion, worse firms loose some ground as compared to the median and best performing ones

Conclusions

- THE NEW PRODUCTION FRAMEWORK GENERATES CENSUS-LIKE SBS DATA THAT NATURALLY LEADS TO **NEW DATA INTEGRATION OPPORTUNITIES** AS WELL AS TO THE INTRODUCTION OF A **NEW CLASS OF INDICATORS**
- NEW **VARIABILITY AND DISTRIBUTION-BASED INDICATORS** CAN BE IMPLEMENTED IN OFFICIAL STATISTICS TO EXPLORE **FIRM LEVEL HETEROGENEITY** THUS PROVIDING ADDITIONAL INFORMATION OF REMARKABLE RELEVANCE FOR BOTH BUSINESS ANALYSTS AND POLICY MAKERS
- **HOW FEASIBLE WOULD BE TO IMPLEMENT THOSE INDICATORS IN YOUR COUNTRY**? CAN WE ADOPT SUB-OPTIMAL SOLUTIONS?