

Serge Allegrezza

Director General of the Institute of Statistics and Economic Studies (STATEC)

Déterminants de la confiance dans la statistique publique. Vers une perspective européenne?

The determinans of trust in official statistics. Heading for a european perspective?

The paper is about measuring trust¹ in official statistics and in the credibility of organizations producing them as public good to support various decision makers and citizens in a democratic society. It has been presented at the 16th Conference of IAOS in September 2018 in Paris (OECD). The long version is under review.

In this paper, the framework proposed by the OECD in 2011 to apprehend trust in official statistics has been adopted as a guideline to draft the questionnaire and organize the data analysis. Trust is now is measured by an agreed set of questions (OECD, 2017) and allows constructing indicators about social capital.

Ignorance of statistical numbers is widely spread; people have a very fuzzy idea of basic macroeconomic indicators, despite being omnipresent and largely commented in the news. In the same context, the OECD has shown in a recent report about adult's skills in literacy and numeracy that 18.5% of adults, on average, have poor reading skills and poor numeracy skills. Around one in four adults has no or only limited experience with computers or lacks confidence in their ability to use computers (OECD, 2016). A recent report on financial literacy shows that in G20 countries, only 52% of adults on average reached a minimum target of basic skills like establishing a budget (OECD 2017). What level of numeracy is required from citizens in order to understand and use official statistics which are deemed essential for running the economy, managing government or intervening in public debate?

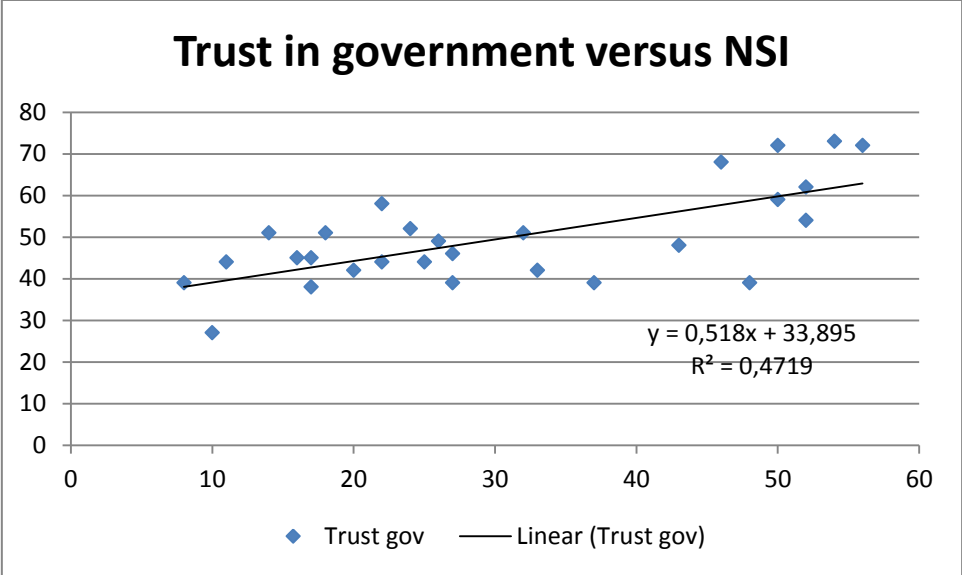
A Eurobarometer poll (2015) has revealed some striking facts. For instance: 6% of Europeans estimated their national growth rates correctly, 31% don't know, most overestimate growth rates. None could give the correct answer on inflation (31% did not know the number, which was highly exaggerated compared to the true number). 23% correctly estimated the unemployment rate in their country, 20% don't know the figure, and all were overestimating the unemployment rate.²

To what extent does this stounding lack of numeracy impact trust in economic statistics? The Eurobarometer survey shows that only half of the interviewees trust economic statistics (10% don't know). There are off course significant differences among countries: it ranks highest in Denmark (83%) and lowest in Cyprus (36%), Luxembourg is 8th (63% expressing trust). I found few analyses, to

¹ Trust and confidence are used interchangeably

² The true number could be questioned, at least in Luxembourg: for example unemployment mixes up the national rate (ADEM) and the harmonized unemployment rate (Eurostat), the same holds for inflation (national price index versus harmonized European). GDP is blurred by the quarterly accounts and successive annual revisions of the same reference year.

my best knowledge, of the relations between trust in official statistics and in NSI's on the one hand and macro-economic factors (economic and social situation, trust in institutions, cultural context...) on the other hand. The same lack of studies, at micro level, exploring relations between trust in numbers and individual characteristics (income, level of education, trust in institutions...) is surprising, given the challenges national statistics are facing.



Trust must be seen in a wider context. For example, using the Eurobarometer data, it is straightforward to depict a linear relation between confidence in official statistics (vertical axes) across Europe and confidence in government (horizontal axes). This means that when interpreting the trust in official numbers or in statistical offices, the more general context has to be taken into account.

Some national statistics institutes collect data on confidence, based on the OECD framework, in official statistics. See for example the studies by New Zealand and United Kingdom. For our study, we will focus on a survey ordered by INSEE and realized by CEVIPOF. Chiche and Chauvrie (2016) resorted to econometric analysis of the defiance by the French public in various statistical fields like growth, unemployment, immigration, public deficit etc. explained by variables on age, gender, education level, income, and political preferences. The results show that trust in official numbers is correlated with confidence in political institutions (government, parliament, president). The results show also differences between gender, age, level of education, income and political preferences. Public statistics are not uniformly appreciated by a heterogonous public. The richer and more educated, the more left leaning, the more interviewed persons tend to trust official statistics. There are also different explanatory patterns depending on the statistical fields.

The Luxembourg surveys

The data

Measurement of trust or confidence (the two terms are used indifferently in our study) in statistics as an institution has been conducted in several representative surveys. The surveys were realized to assess the reputation of the Institute, to foster better products and to improve targeted communication. Besides the "peer review", based on the European code of conduct of official

statistics, which aims at assessing the conformity with agreed quality standards, the surveys presented in this study look for the *perception* of statistics by a representative sample of persons. As official statistics and public numbers should be as objective as possible, the professional reputation of the national institutes is of utmost importance. The questions aim at measuring the degree of confidence or trustworthiness of statistical numbers and the confidence in the national statistical institute (STATEC).

The explanatory variables are of two sorts. First, there are variables on confidence. The most contentious is the perceived political independence of the Institute (Independence). The confidence in the handling of personal data (Protection) is very important, given the sensitivity of the public on this issue. The confidence in Media (Media) is important since it represents still the most important channel to spread statistical data, press statements and studies. There are two behavioural variables. The first derives from a question on whether respondents use or refer to statistics published by the Institute (Usage), which refers to their behaviour, validating the trust in official numbers. The second derives from a question asking whether the interviewee participated in a survey realized by the national institute. This variable shows to what extent contributors to an official survey (LFS, EUSILC ...), who give some of their precious time to respond to questions, trust the numbers and the Institute (Participant).

Second, the surveys collected data on socio-demographic characteristics. There are three variables: income categories (Income), education levels (Education), gender (sec) and age categories (age). The survey of 2015 includes also some questions about economic opinions, used as proxy for political preferences. Questions appreciating public debt level and the public deficit were meant to capture typical attitudes of fiscal conservatives versus fiscal progressives.

Some remarks on the expected results: as statistics should be a public good, confidence should be non-discriminatory: no gender and “class differences” by income and education, independence from government and from political interferences. So the best econometric result would consist in no significant effects from those variables on confidentiality. In sharp contrast, participation in surveys and usage of official statistics should have positive, strong and significant effects on trust in statistics. Those are the ideal results we are aiming at.

The survey of 2017 includes two other questions about trust in other fellow citizens, family and relatives, neighbours and a question about political preferences (left- right scale).

Results of the analysis

A bunch of various specifications of logit regressions have been run taking as dependent variables: trust in public statistics, political independence, use of statistics by citizens and institutions and a positive opinion on the performance of the NSI (STATEC). The dependent variables of the four equations have been used as determinants (exogenous variables) for the other regressions.

The 2015 survey

Confidence in public statistics is influenced positively by the perception of political independence as well as guaranteeing personal data protection, which is highly and positively significant. Those are the expected effects and reassuring results.

The confidence in media, used as main distribution mechanism of statistics, is positively linked to trust in statistics. This is important since news reporting is often biased or contaminated by

comments or criticism (online comments on social media). Participating in an almost compelling survey organized by STATEC has no significant impact on trust in official statistics.

The level of education has a positive impact on trust in statistics. In contrast, gender (man) and income (dummies) have no significant effect. Trust in statistics diminishes with age. Those comments hold for the results of the equation on trust in the NSI (STATEC).

Trust in statistics has a significant positive impact on the perception of political independence. Interestingly, the 2015 survey shows a negative impact of nationality (Luxembourger) on political independence. Nationals are less convinced by the political independence of official numbers but show no effect on trust in the NSI.

Another interesting result is the use of statistics which is positively impacted by participating in an official survey. This suggests that interviewees are more prone to use statistics when they are becoming aware of them by giving data to the NSI.

The extended model included two additional sets of variables on knowledge of basic statistics (estimate of the public debt, percentage of GDP, unemployment rate, ...) and two opinion questions on whether the government is entitled to run a public deficit during recessions and whether public debt is always excessive.

Data shows that ignorance of public data is widely spread (less than half of respondents gave the right answer, except for unemployment where more than half gave the right answer) and confidence works as a substitute for some sort of minimum knowledge. The variables did not turn out as significant and no additional explanation emerged. The only variable with a significant (negative) impact on trust in official statistics is the variable on fiscal conservatism (55% of the sample hold the opinion that public debt is always excessive, also it is less than 25% of gdp, the second lowest in the euro area). But no effect on other dependent variables has been detected. It suggests that trust can be influenced by political opinions.

Some other techniques have been used to take care of endogeneity. Some of the dependent variables could be exogenous in the other equations of interest in a simultaneous equation framework. To take care of this effect, simultaneous equations models for dichotomous dependent variables, bivariate and trivariate probits³ have been used. For example trust in statistics and in the NSI might have contemporaneous mutual influence asking for a more sophisticated model. But the results (reported in a companion document, available on request with the Stata code) did not show any significant difference to the results reported in table 1.

The 2017 survey

The same regressions have been run on the new data gathered from a survey realized in 2017 with the same questionnaire. Table 2 summarizes the results for the four equations: trust in official statistics, political independence, usage of statistics and trust in the NSI (STATEC).

Trust in official statistics is determined by trust in the institute, the political independence; protection of personal data and the usage of statistics as well as media confidence have all a positive and significant effect. Education and age have a significant negative impact. To appreciate the magnitude of the effects, it is interesting to have a look at marginal effects, i.e. the probability of

³ SEM and bi and trivariate probit in STATA

trusting given the individual determinants taken one by one. The most important effect in size is trust in the NSI (STATEC) followed by the level of education, then come political independence and data protection (cf. table in the appendix).

Table 1. Logit regressions on trust, political independence, usage and trust in NSI (Survey of 2017)

Survey 2017	Trust in Public Statistics	Political Independence	Usage of statistics	Trust in NSI (STATEC)
Trust	x	+++	ns	+++
Trust in NSI(STATEC)	+++	+++	+++	x
Independence Political	+++	x	ns	+++
Usage of statistics	++	ns	x	+++
Protection of data	++	+++	+++	+++
Participant Official survey	ns	+	+++	++
Media trust	+++	+++	ns	+++
Sex (male)	ns	+	+++	ns
Income (categories)	ns	ns	---	ns
Education (levels)	+++	+	+	ns
Age (categories)	---	+	---	+
National	ns	ns	+	ns
McFadden R2 N=2589	0.229	0.213	0.097	0.384

Significance $p < 0.001$: +++ or --- $p < 0,01$ ++ or -- $p < 0,05$: + or - NS = not significant X= excluded

The additional variables on political attitudes (left to right) and confidence in other fellow citizens have been added to the basic regressions. (In the appendix, the full regressions results are exhibited for trust in statistics). It turns out that political preference (dummy variable coded 1 for left) have no significant effect while trust in fellow citizens (variable altrui) has a significant and positive effect in all the specifications. Confidence in fellow citizens, a dimension of social capital, increases the probability to trust statistical figures from the NSI.

The results of the surveys of carried out in 2017 and 2015, show very similar results.

Conclusions

The two data sets on trust in statistics used in this study are quite rich and provided some useful information, confirming the “felt” political independence of official statistics.

The rate of trust in official statistics reaches circa 69% in 2017, which is quite similar to what its value in 2015 and before (at par with trust in police). We showed that political independence reinforces confidence but political preferences have no bearing on confidence in statistics. There are still more than 30% of respondents which have only limited confidence in official statistics! Which represents a challenge to communication? This raises the question of the optimal level of trust at a given point in time and in a particular social and economic context.

Some methodological considerations are unavoidable .The analysis has to be carried forward to extract all the relevant information. Multiple questions on trust in institutions and fellow citizens should be explored furthermore using factor analysis. Logit regressions with ordinal variables should also be completed to take into account a greater variability of the responses (full trust, partial trust, low trust and no trust at all). Ordinal models should be tested as a consequence. The samples which are weighted to increase representativeness should be critically assessed in order to correct for additional biases arising with online surveys. Finally, the two samples of 2015 and 2017 could be pooled and used for further analysis.