Empirical Examples of Big Internet Data for Macroeconomic Nowcasting [1][2]

Fotis Papailias

King's College London

CARMA – 13 July 2018

[1] Based on joint work with George Kapetanios, Massimiliano Marcellino, Gian Luigi Mazzi.

[2] Part of this presentation is based on a report prepared for Eurostat, that we thank for financial support and many useful comments and suggestions.

Outline

Outline

1. **Introduction**
   - Types of Big Data
   - Pros and Cons of Big Data for Macroeconomic Nowcasting
   - Data & Methodological Issues
2. **Summary of Literature Review**
   - Existing literature on the use of Big Data
   - Existing literature on Macroeconomic Nowcasting
   - What about their combination?
3. **Summary of Big Data Modelling**

Outline

4. **Empirical Analysis**
   - Data Description
   - Nowcasting & Short-Term Forecasting Evaluation
5. **Conclusions and Policy Recommendations**

Introduction

Introduction: Big Data Issues

Data Availability. Most data pass through private providers and are related to personal aspects.

Hence, continuity of data provision could not be guaranteed.

For example, Google could stop providing Google Trends, or at least no longer make them available for free.

Or online retail stores could forbid access to their websites to crawlers for automatic price collection.

Or individuals could extend the use of softwares that prevent tracking their internet activities, or tracking could be more tightly regulated by law for privacy reasons.

Introduction: Big Data Issues

Digital Divide. The fact that a sizable fraction of the population still has no or limited internet access.

This implies that the available data are subject to a sample selection bias, and this can matter for their use.

Suppose, for example, that we want to nowcast unemployment at a disaggregate level, either by age or by regions.

Internet data relative to older people or people resident in poorer regions could lead to underestimation of their unemployment level, as they have relatively little access to internet based search tools.

Introduction: Big Data Issues

Changing Size & Quality. A third issue is that both the size and the quality of internet data keeps changing over time, in general much faster than for standard data collection.

For example, applications such as Twitter or WhatsApp were not available just a few years ago, and the number of their users increased exponentially, in particular in the first period after their introduction.

Similarly, other applications can be gradually dismissed or used for different uses. For example, the fraction of goods sold by EBay through proper auctions is progressively declining over time, being replaced by other price formation mechanisms.

Introduction: Big Data Issues

Bias in Answers. Again more relevant for digital than standard data collection, is that individuals or businesses could not report truthfully their experiences, assessments and opinions.

For example, some newspapers and other sites conduct online surveys about the feelings of their readers (happy, tired, angry, etc.) and one could think of using them, for example, to predict election outcomes, as a large fraction of happy people should be good for the ruling political party.

But, if respondents are biased, the prediction could be also biased, and a large fraction of non-respondents could lead to substantial uncertainty.

Introduction: Big Data Issues

Data Format. A fifth issue is that data could not be available in a numerical format, or not in a directly usable numerical format.

A similar issue emerges with standard surveys, for example on economic conditions, where discrete answers from a large number of respondents have to be somewhat summarized and transformed into a continuous index.

However, the problem is more common and relevant with internet data.

Introduction: Big Data Issues

Irregularities. A final issue, again common also with standard data but more pervasive in internet data due to their high sampling frequency and broad collection set, relates to data irregularities:

- outliers,
- working days effects,
- missing observations,
- presence of seasonal / periodic patterns, etc.

all of which require properly de-noising and smoothing the data.

Introduction: Big Data Advantages

Big data provide potentially relevant *complementary* information with respect to standard data, being based on rather different information sets.

Moreover, they are timely available and, generally, they are not subject to subsequent revisions, all relevant features for potential coincident and leading indicators of economic activity.

Finally, they could be helpful to provide a more granular perspective on the indicator of interest, both in the temporal and in the cross-sectional dimensions.

Introduction: Big Data Advantages

In the temporal dimension, they can used to update nowcasts at a given frequency, such as weekly or even daily, so that the policy and decision makers can promptly update their actions according to the new and more precise estimates.

In the cross-sectional dimension, big data could provide relevant information on units, such as regions or sectors, not fully covered by traditional coincident and leading indicators.

Introduction: Summary

First, do we get any relevant insights? In other words, can we improve nowcast precision by using big data?

Second, do we get a big data hubris? Again as anticipated, we think of big data based indicators as complements to existing soft and hard data-based indicators, and therefore we do not get a big data hubris.

Third, do we risk false positives? Namely, can we get some big data based indicators that nowcast well just due to data snooping?

Introduction: Issues Summary

Fourth, do we mistake correlations for causes?

Fifth, do we use the proper econometric methods?

Sixth, do we have instability due to Algorithm Dynamics or other causes (e.g., the financial crisis, more general institutional changes, the increasing use of internet, discontinuity in data provision, etc.)?

Finally, do we allow for variable and model uncertainty?

Literature Review: Main Findings

The purpose of the literature review is to answer four questions:

1. What are possible big data sources in relation to macroeconomic indicators?
2. What are the advantages and disadvantages for each of the previously analysed sources?
3. What are the main types of statistical methods used in the big data in macroeconomics literature?
4. What are the possible gains generated either by the use of big data or new statistical methods or both in comparison with existing practices in the field of nowcasting?

Literature Review: Main Findings

As anticipated in the Introduction, after a careful examination of the most important papers in each area we can say that the majority of big data papers are based on Google Trends as predictors.

The advantages of using data like Google Trends include:

- **a** the improved timeliness of the forecasts without need for data revision,
- **b** the potential improvement of forecasts,
- **c** open access to the data,
- **d** easy data handling,
- **e** good data quality,
- **f** reasonable possibility that this sort of data will be released on a continuous basis.

Literature Review: Main Findings

However, one must be cautious when using data of this type as it is often associated with the following issues:

- **a** the use of Google data as the only data input could lead to biased results (commonly known as "big data hubris"),
- **b** the restrictions to access the raw data but only the Google index,
- **c** the possibility that free access will be discontinued.

Therefore, we suggest the use of Google Trends for nowcasting the macroeconomic variables of interest for Eurostat.

We strongly believe that such data must be used as a supplement to current forecasting tools and not as a substitute.

Literature Review: Main Findings

Regarding questions (3) and (4), it turns out that most of the papers in the literature generate nowcasts based on mixed frequency versions of linear regressions, VARs, (dynamic) factor models or a combination of them, and adopt various strategies for variable selection in the presence of a large set of potential regressors.

While no clear-cut ranking of the alternative methodologies emerge, there seems to be consensus about the usefulness of big data for nowcasting variables such as unemployment, GDP, inflation and surveys, even though the gains are often computed with respect to (too) simple benchmarks.

Part 2: Big Data Modelling

Big Data Modelling

Let y_t, $t = 1, ..., T$, be the target variable and $x_t = (x_{1t}, ..., x_{Nt})'$ be a set of potential predictors, with N very large.

We do not assume a particular data generating process for y_t but simply posit the existence of a representation of the form

$$y_t = a + g(x_{1t}, ..., x_{Nt}) + u_t, \quad (1)$$

which implies that $E(u_t | x_{1t}, ..., x_{Nt}) = 0$.

While the potential nonlinearity in (1) might, in principle, be worth exploring, it is extremely difficult to model nonlinearities in the context of large datasets and no work is available on this in the big data literature.

Big Data Modelling

As a result, we consider an approximating linear representation of the form,

$$y_t = a + \sum_{i=1}^{N} \beta_i x_{it} + u_t, \tag{2}$$

with u_t denoting a martingale difference process and where the set of x_{it}s can also contain products of the original indicators in order to provide a better approximation to (1).

Big Data Modelling

Our main aim is to provide estimates for current and future values of y_t, where either no, or only a preliminary, value for y_t is available from official statistics.

To do so, we can rely on many approaches, which can be categorised in three main strands.

The **first strand** aims to provide estimates for $\beta = (\beta_1, ..., \beta_N)'$.

While ordinary least squares (OLS) is the benchmark method for doing so, it is clear that if N is large this is not optimal or even feasible (when $N > T$).

Big Data Modelling

So other methods need to be used. We consider two classes of methods.

- The first one is sparse regression, with origins in the machine learning literature. A main aim there is to stabilise the variability of the estimated β_i.
- The second class considers the use of a variety of information criteria such as AIC or BIC to select a smaller subset of all the available predictors.

Big Data Modelling

The **second strand** consists of reducing the dimension of x_t by producing a much smaller set of generated regressors, which can then be used to produce nowcasts and forecasts in standard ways.

The **third strand** suggests the use of a (possibly very large) set of small models, one for each available indicator or small subset of them, and then the combination of the resulting many nowcasts or forecasts.

Big Data Modelling: Main Findings

We feel that Multiple Testing (related to Boosting) and, potentially, some variant of LASSO could be the preferred approaches among the set of machine learning techniques.

Among the data reduction techniques, PCA and, possibly PLS are promising.

And it would be also worthwhile experimenting with Bayesian regression, with substantial shrinkage, and forecast combination, with simple equal weighting.

Big Data Modelling: Main Findings

Finally, all these approaches should be also modified to take into account the possibility of a different timing for the target and the indicator variables.

We have surveyed a number of alternative methods to handle mixed frequencies, and it turns out that Unrestricted MIDAS or bridge modelling appear as the most promising approaches, as they preserve linearity and do not add an additional layer of computational complexity.

Part 3: Empirical Analysis

Empirical Analysis: Introduction

We consider nowcasting and short-term (one- to twelve- months ahead) forecasting three key macroeconomic variables:

1. **inflation** (measured by the growth rate in the Consumer Price Index),
2. growth in **retail sales** (measured by the growth rate of the Retail Trade Index), and
3. the **Unemployment Rate**.

The exercise is conducted recursively in a pseudo out of sample framework, using monthly data for three economies:

- Germany (DE),
- Italy (IT) and
- the UK.

Empirical Analysis: Introduction

We assess the relative performance of:

- Big Data (proxied via weekly Google Trends) and
- standard indicators (based on a large set of weekly and monthly economic and financial variables)

In fact, as we have mentioned several times, we think of Big Data as providing complementary information, and we wish to assess how useful it is in a forecasting context relative to standard indicators.

Empirical Analysis: Introduction

We also evaluate the role of several econometric methods and alternative specifications for each of them (with or without big data), also capable of handling the frequency mismatch in our data. Specifically, we consider:

- Naive and autoregressive (AR) models as benchmarks
- Dynamic Factor Analysis (DFA)
- Partial Least Squares (PLS)
- Bayesian Regression (BR) and
- LASSO regression.

DFA and PLS are representatives of data reduction methods; BR and LASSO are representatives of, respectively, econometric and machine learning techniques.
In addition, we also consider model averaging. Overall, we have a total of 255 models and model combinations.

Germany, h=0					
CPI		**RS**		**UN**	
Naive	1.455	**Naive**	1.415	**Naive**	1.342
AR(1)-GFac	0.985	**AR(1)-GFacUMIDAS**	1.002	**AR(1)-Gall**	0.990
Ave12	0.949	**Ave24**	0.869	**Ave6**	0.974
DFA2(ave+GUMIDAS+AR1)	1.023	**DFA4(ave+Gall)**	0.901	**DFA2(ave+Gall)+AR(1)**	0.985
DFA2(ave)	1.028	**DFA4(ave)**	0.899	**DFA4(ave)+AR(1)**	0.979
PLS1(umidas+Gall)+AR(1)	1.075	**PLS1(ave)+AR(1)+Gall**	1.098	**PLS1(ave)+AR(1)+Gall**	1.003
PLS1(umidas)+AR(1)	1.076	**PLS1(ave)**	1.113	**PLS1(ave)+AR(1)**	1.011
BR(0.5)(ave+Gall+AR1)	0.962	**BR(0.5)(umidas+GUMIDAS+AR1)**	0.883	**BR(0.5)(ave)+AR(1)+Gall**	0.982
BR(0.5)(ave)+AR(1)	0.978	**BR(1)(umidas)**	0.887	**BR(1)(ave)+AR(1)**	0.998
LASSO(umidas+Gall+AR1)	1.060	**LASSO(ave+Gall)**	0.980	**LASSO(umidas)+AR(1)+GaveUMIDAS**	0.978
LASSO(ave)	1.058	**LASSO(ave)**	0.985	**LASSO(ave)+AR(1)**	0.993
Best5.1.GOOG	1.064	**Best5.12.ALL**	0.921	**Best1.12.ALL**	1.006
Best5.1.NOGOOG	1.020	**Best3.12.NOGOOG**	0.902	**Best5.1.NOGOOG**	1.110

Germany, h=1					
CPI		RS		UN	
Naive	1.503	Naive	1.648	Naive	1.227
AR(1)-GFac	1.015	AR(1)-GaveUMIDAS	1.000	AR(1)-GFacUMIDAS	0.998
Ave12	0.951	Ave24	0.980	AR(1)	1.000
DFA4(ave+Gall)+AR(1)	1.010	DFA3(umidas+Gall+AR1)	0.957	DFA2(umidas+Gall)+AR(1)	1.030
DFA4(ave)+AR(1)	1.010	DFA3(umidas)	0.956	DFA2(umidas)+AR(1)	1.030
PLS2(umidas+Gall+AR1)	0.961	PLS4(ave+Gall)	0.924	PLS2(umidas+Gall)+AR(1)	1.052
PLS2(umidas)+AR(1)	0.963	PLS1(ave)+AR(1)	0.924	PLS2(umidas)+AR(1)	1.049
BR(0.5)(ave+GUMIDAS)+AR(1)	0.975	BR(1)(umidas+GUMIDAS+AR1)	1.014	BR(0.5)(umidas+Gall)+AR(1)	1.001
BR(1)(umidas)+AR(1)	1.001	BR(1)(umidas)+AR(1)	1.027	BR(0.5)(umidas)+AR(1)	1.001
LASSO(umidas+Gall)	1.046	LASSO(umidas)+AR(1)+Gave	0.906	LASSO(ave+GUMIDAS)	0.976
LASSO(umidas)+AR(1)	1.039	LASSO(umidas)+AR(1)	0.913	LASSO(ave)	0.982
Best1.3.GOOG	0.982	Best3.12.ALL	0.925	Best1.3.ALL	1.021
Best5.1.NOGOOG	0.938	Best5.6.NOGOOG	0.948	Best5.6.NOGOOG	1.109

Italy, h=0					
CPI		**RS**		**UN**	
Naive	1.281	**Naive**	1.158	**Naive**	1.495
AR(1)-GFacUMIDAS	0.979	**AR(1)-Gall**	0.986	**AR(1)-Gave**	1.000
Ave6	0.718	**Ave24**	0.670	**AR(1)**	1.000
DFA2(ave+GUMIDAS)	0.769	**DFA2(ave+GUMIDAS)**	0.742	**DFA2(umidas+GUMIDAS)**	0.982
DFA2(ave)	0.784	**DFA2(ave)**	0.745	**DFA2(umidas)**	0.993
PLS1(umidas+Gall)	0.956	**PLS1(ave+Gall)**	0.996	**PLS1(umidas+Gall)**	1.182
PLS1(umidas)	0.951	**PLS1(ave)**	0.980	**PLS1(umidas)**	1.181
BR(1)(umidas+Gall)	0.781	**BR(1)(umidas+GUMIDAS+AR1)**	0.720	**BR(1)(umidas+GUMIDAS)**	0.985
BR(1)(umidas)	0.781	**BR(1)(umidas)**	0.734	**BR(1)(umidas)**	0.986
LASSO(umidas+Gall)	0.902	**LASSO(ave+GUMIDAS)**	0.947	**LASSO(umidas+GUMIDAS+AR1)**	1.024
LASSO(umidas)	0.876	**LASSO(ave)**	0.966	**LASSO(umidas)**	1.030
Best5.12.ALL	0.752	**Best3.12.ALL**	0.549	**Best5.12.ALL**	0.911
Best1.12.NOGOOG	0.742	**Best1.12.NOGOOG**	0.585	**Best1.12.NOGOOG**	0.912

Italy, h=1					
CPI		RS		UN	
Naive	1.035	**Naive**	1.410	**Naive**	1.241
AR(1)-GFac	1.019	**AR(1)-Gave**	1.001	**AR(1)-Gall**	0.965
Ave3	0.832	**Ave12**	0.910	**Ave24**	0.957
DFA2(umidas+GUMIDAS+AR1)	0.966	**DFA2(umidas+GUMIDAS+AR1)**	0.997	**DFA2(ave)+AR(1)+Gall**	0.913
DFA3(ave)	0.967	**DFA2(umidas)**	0.997	**DFA2(ave)+AR(1)**	0.953
PLS1(umidas+Gall)	0.970	**PLS1(umidas+Gall+AR1)**	1.028	**PLS3(umidas)+AR(1)+Gall**	0.927
PLS1(umidas)	0.968	**PLS1(umidas)**	1.027	**PLS2(umidas)**	0.953
BR(0.5)(ave+Gall+AR1)	0.928	**BR(1)(umidas+Gall)+AR(1)**	0.997	**BR(1)(umidas)+AR(1)+Gall**	0.964
BR(0.5)(umidas)	0.929	**BR(1)(umidas)+AR(1)**	0.996	**BR(1)(ave)**	0.990
LASSO(umidas+GUMIDAS+AR1)	0.960	**LASSO(umidas)+AR(1)+Gave**	0.994	**LASSO(ave)+AR(1)+Gall**	0.964
LASSO(umidas)	0.991	**LASSO(umidas)**	0.994	**LASSO(umidas)**	0.989
Best3.12.ALL	0.878	**Best1.12.ALL**	0.753	**Best3.12.ALL**	0.849
Best3.12.NOGOOG	0.959	**Best3.12.NOGOOG**	0.787	**Best5.12.NOGOOG**	0.888

UK, h=0					
CPI		**RS**		**UN**	
Naive	1.444	**Naive**	1.452	**Naive**	0.881
AR(1)-Gave	0.994	**AR(1)-Gave**	0.993	**AR(1)-Gall**	0.982
Ave24	0.917	**Ave24**	0.860	**Ave3**	0.877
DFA3(ave+GUMIDAS)	0.932	**DFA3(ave+GUMIDAS)**	0.872	**DFA2(umidas)+AR(1)+Gall**	0.988
DFA3(ave)	0.936	**DFA2(ave)**	0.878	**DFA2(umidas)+AR(1)**	1.003
PLS1(umidas+Gall)	1.238	**PLS1(umidas+Gall)**	0.951	**PLS4(umidas+Gall+AR1)**	0.911
PLS1(umidas)	1.233	**PLS1(umidas)**	0.950	**PLS4(umidas)+AR(1)**	0.921
BR(1)(umidas+GUMIDAS+AR1)	0.984	**BR(0.5)(ave+Gall)**	0.835	**BR(0.5)(ave)+AR(1)+Gall**	0.956
BR(1)(umidas)	0.986	**BR(0.5)(ave)**	0.838	**BR(0.5)(ave)+AR(1)**	0.972
LASSO(umidas+Gall+AR1)	1.318	**LASSO(umidas+Gall)**	0.862	**LASSO(umidas)+AR(1)+Gall**	0.932
LASSO(umidas)	1.319	**LASSO(umidas)**	0.868	**LASSO(umidas)+AR(1)**	0.945
Best1.6.ALL	0.902	**Best5.6.GOOG**	0.947	**Best1.1.ALL**	0.844
Best1.6.NOGOOG	0.949	**Best3.6.NOGOOG**	0.913	**Best1.3.NOGOOG**	0.971

UK, h=1					
CPI		**RS**		**UN**	
Naive	1.279	**Naive**	1.412	**Naive**	0.870
AR(1)-GFac	1.048	**AR(1)-GFac**	1.004	**AR(1)-Gave**	0.999
Ave24	0.936	**Ave24**	0.931	**Ave3**	0.872
DFA2(ave+Gall)	0.972	**DFA3(umidas+Gall)**	0.988	**DFA2(ave+GUMIDAS)+AR(1)**	0.999
DFA2(ave)	0.973	**DFA3(umidas)**	0.988	**DFA2(ave)+AR(1)**	1.000
PLS1(umidas+Gall)	0.986	**PLS3(ave+Gall)**	0.925	**PLS3(umidas)+AR(1)+Gall**	0.978
PLS1(umidas)	0.987	**PLS3(ave)**	0.935	**PLS3(umidas)+AR(1)**	0.979
BR(0.5)(umidas+Gall)	0.978	**BR(0.5)(ave+Gall)**	0.988	**BR(0.5)(ave)+AR(1)+Gave**	0.993
BR(0.5)(umidas)	0.978	**BR(0.5)(ave)**	0.987	**BR(0.5)(ave)+AR(1)**	0.993
LASSO(umidas+Gall+AR1)	1.044	**LASSO(umidas+Gall)**	0.956	**LASSO(ave)+AR(1)+GaveUMIDAS**	0.987
LASSO(umidas)	1.029	**LASSO(umidas)**	0.953	**LASSO(ave)+AR(1)**	0.987
Best1.12.GOOG	0.959	**Best3.12.ALL**	1.003	**Best1.1.ALL**	0.836
Best5.12.NOGOOG	0.946	**Best5.1.NOGOOG**	0.995	**Best5.6.NOGOOG**	0.977

Empirical Analysis: Summary results

	MAE	CPI-DE	RS-DE	UN-DE	CPI-IT	RS-IT	UN-IT	CPI-UK	RS-UK	UN-UK
h=0	**Best Google**	0.962	0.883	0.978	0.752	0.549	0.911	0.902	0.835	0.844
	Best Non-Google	0.949	0.869	0.974	0.718	0.585	0.912	0.917	0.838	0.877
	Diff in %	-1.392%	-1.652%	-0.392%	-4.712%	**6.092%**	**0.186%**	**1.699%**	**0.400%**	**3.868%**
h=1	**Best Google**	0.961	0.906	0.976	0.819	0.753	0.854	0.959	0.925	0.836
	Best Non-Google	0.938	0.913	0.982	0.796	0.787	0.888	0.936	0.931	0.870
	Diff in %	-2.411%	**0.747%**	**0.603%**	-2.893%	**4.352%**	**3.853%**	-2.477%	**0.553%**	**3.816%**

Data-Driven Automated Strategy

- IT CPI: 100% of the times the chosen models include Google Trends,
- IT RTI: 35.14% of the times the chosen models include Google Trends,
- IT UN: 100% of the times the chosen models include Google Trends,
- UK CPI: 53.49% of the times the chosen models include Google Trends,
- UK UN 35.42% of the times the chosen models include Google Trends.

Conclusions

Overall Recommendations

Overall, our suggestion is to take a pragmatic approach that balances potential gains and costs from the use of Big Data for nowcasting macroeconomic indicators, **in addition to** standard indicators.

A preliminary step should be an a priori assessment of the potential usefulness of Big Data for a specific indicator of interest, such as GDP growth, inflation or unemployment.

Overall Recommendations

This requires to evaluate the quality of the existing nowcasts and whether any identified problems, such as bias or inefficiency or large errors in specific periods, can be fixed by adding information as potentially available in Big Data based indicators.

Similarly, it should be considered whether these additional indicators could improve the timeliness, frequency of release and extent of revision of the nowcasts.

Relevant information can be gathered by looking at existing empirical studies focusing on similar variables or countries, and in this respect the extensive literature review we presented can be quite helpful.

Overall Recommendations

Once Big Data passes the "need check" in the preliminary step, **the first proper step** of the Big Data based nowcasting exercise is a careful search for the specific Big Data to be collected.

As we have seen, there are many potential providers, which can be grouped into Social Networks, Traditional Business Systems, and the Internet of Things.

Naturally, it is not possible to give general guidelines on a preferred data source, as its choice is heavily dependent on the target indicator of the nowcasting exercise.

Overall Recommendations

Having identified the preferred source of Big Data, **the second step** requires to assess the availability and quality of the data.

A relevant issue is whether direct data collection is needed, which can be very costly, or a provider makes the data available.

In case a provider is available, its reliability (and cost) should be assessed, together with the availability of meta data, the likelihood that continuity of data provision is guaranteed, and the possibility of customization (e.g., make the data available at higher frequency, with a particular disaggregation, for a longer sample, etc.).

Overall Recommendations

All these aspects are particularly relevant in the context of applications in official statistical offices.

As the specific goal is nowcasting, it should be also carefully checked that the temporal dimension of the Big Data is long and homogeneous enough to allow for proper model estimation and evaluation of the resulting nowcasts.

Overall Recommendations

The third step analyzes specific features of the collected Big Data.

A first issue that is sometimes neglected is the amount of the required storage space and the associated need of specific hardware and software for storing and handling the Big Data.

A second issue is the type of the Big Data, as it is often unstructured and may require a transformation into cross-sectional or time series observations.

Overall Recommendations

Even when already available in numerical format, pre-treatment of the Big Data is often needed to remove deterministic patterns and deal with data irregularities, such as outliers and missing observations.

While standard methods can be usually applied, the size of the datasets suggests to resort to robust and computationally simple approaches, applied variable by variable.

Overall Recommendations

The fourth step requires to assess the presence of a possible bias in the answers provided by the Big Data, due to the "digital divide" or the tendency of individuals and businesses not to report truthfully their experiences, assessments and opinions.

A related problem, particularly relevant for nowcasting, is the possible instability of the relationship with the target variable.

This is a common problem also with standard indicators, as the type and size of economic shocks that hit the economy vary over time. Both issues can be however tackled at the modelling and evaluation stages.

Overall Recommendations

The fifth step when nowcasting with Big Data requires to select the proper econometric technique.

Here, it is important to be systematic about the correspondence between the nature of the Big Data setting and use under investigation and the method that is used.

There is a number of dimensions along which we wish to differentiate.

Overall Recommendations

To conclude, we are very confident that Big Data are precious also in a nowcasting context, not only to reduce the errors but also to improve the timeliness, frequency of release and extent of revision.

We hope that the approach we have developed in this presentation will be useful for many users.