BIG DATA FOR MEASURING THE INFORMATION SOCIETY

COUNTRY REPORT
SWEDEN

Acknowledgments:

The big data analysis contained in this report was carried out by Ivan Vallejo Vall, Market Analyst at the ICT Data and Statistics Division within the Telecommunication Development Bureau of ITU. Fredrik Eriksson, Statistician in the same division, made substantial contributions to the analysis.

The Internet Foundation in Sweden (IIS) provided the data used in this report, as well as insightful information on the methodology of the data collection. IIS's cooperation is duly acknowledged.

CONTENTS

1. BACKGROUND AND CONTEXT

1.1. PROJECT DESCRIPTION

This aim of the project is to use big data from the telecom industry to improve and complement existing statistics and methodologies to measure the information society. The results of this project are expected to help countries and ITU to produce official ICT statistics and to develop new methodologies by combining new and existing data sources. This will directly benefit policy makers who will have access to new official statistics and benchmarks. It will also benefit data producers at the national level by guiding them in the use of big data for ICT measurement.

Specifically, the project:

- Combines the role of ITU as a standard-setting organisation in terms of global ICT measurement with official data producers' experience and interest in working with big data. Those data producers who can and are willing to share big data are invited to join the project to explore new methodologies and data sources to produce official ICT statistics. ICT data producers include: national statistical offices, telecommunication regulatory authorities, telecommunication service providers (Mobile Network Operators – MNOs), and Internet Service Providers (ISP).

- Takes advantage of ITU's involvement and experience in existing big data initiatives: ITU has organized a number of big data sessions in terms of regulation and monitoring, published information on the opportunities and challenges of big data, and developed a big data project on mobile phone data for health. ITU is also an active member of the UN Global Working Group on Big Data for Official Statistics, has networked with public and private organizations working on the topic, and has identified a number of potential (public and private) partners that would be interested and benefit from participating in this project.

- Explores and analyses the kind of new information society data and statistics that ITU and other stakeholders (including policy makers, analysts and other data producers and users) would value most, and which are currently not existing.

- Engages with ICT data producers to discuss the development of specific indicators for new statistics that could be produced through big data analytics, and by combing new and existing data sources and methodologies.

1.2. PILOT COUNTRY CONTEXT

The pilot in Sweden is of different nature compared to the other pilot countries as the focus is on analyzing big data on broadband speed measurements rather than data obtained from mobile operators.

From a consumer perspective, broadband speed is one of the most important factors when choosing an Internet connection. For instance, in the European Union (EU) the download speed is the second most cited factor, after price, when deciding which broadband plan to choose (TNS Opinion & Social, 2014).[1] However, six out of ten EU citizens do not know the maximum download speed of their broadband Internet plan and, among those that know it, a quarter of them believe that their real speed does not correspond to the one specified in their contract. Moreover, four in ten households in the EU admit having experienced difficulties accessing content at home because of speed or capacity issues (TNS Opinion & Social, 2014). It can thus be concluded that Internet speed is both an important and a controversial factor for consumers.

It is also one of the parameters often monitored to ensure that telecommunication operators and Internet service providers (ISPs) comply with some minimum quality-of-service (QoS) requirements. Monitoring quality of service is a key component of the mandate of national regulatory authorities. For example, the Telecom Regulatory Authority of India sets that subscribers should get a minimum of 80 per cent of the speed specified in their contract, as measured from the ISP node to the user (Telecom Regulatory Authority of India, 2006).

From a policy perspective, broadband speeds also have wide implications concerning the initiatives undertaken in the telecommunication sector. For instance, the definition of broadband is often tied to a given minimum speed. Broadband speeds may also be an important determinant of broadband impact on economic growth.

The Internet Foundation in Sweden (IIS) is an independent public-service organization responsible for the operation of the top-level domains '.se' and '.nu'.[2] IIS reinvests part of the revenues obtained from the administration of these domains in , inter alia, the promotion of research on the Internet. IIS also provides a consumer tool *Bredbandskollen* ("the Broadband Check"), a software-based Internet platform to measure actual broadband speeds, which helps consumers easily test and evaluate their internet connection, for both fixes and mobile subscriptions.[3] The tool was launched in 2007 and IIS has 10 years of user-generated data from this tool.

Bredbandskollen measures the download and upload speed by sending data to and receiving data from the closest Internet exchange point (IXP) to the end user. Currently, IIS has test servers hosted in five IXPs (in Malmö, Göteborg, Stockholm, Sundsvall and Oslo), against which the tests are run. As a result, measurements often take place in an IXP with which the major ISPs in Sweden will have a direct interconnection. The download and upload speed tests last ten seconds each. During the test, average speeds are calculated for the first two seconds and for the entire ten seconds. The highest of the two results is retained. The tool also measures latency, i.e. the delay or round-trip time to the closest server, and has been optimized to measure broadband speeds in Sweden and in the neighboring border areas.

IIS annually produces a report with key aggregates on broadband speeds across regions and operators; however, additional analysis could provide more insights with respect to other international

---

[1] On average, 41 per cent of respondents in the EU mentioned download speed as an important factor when subscribing to an Internet connection, compared with 71 per cent of respondents citing price. The figures refer to fieldwork carried out in January 2014.

[2] For more information, see https://www.iis.se/english/

[3] For more information about Bredbandskollen, see http://www.bredbandskollen.se/reports

classificatory variables, e.g. urban and rural. The broadband speed data could also complement and be compared against other broadband speed measurement platforms, e.g. Ookla, Akamai and SamKnows, as well as figures reported to ITU on advertised speed for fixed broadband subscriptions.

1.3. STAKEHOLDERS IN THE PILOT AND PROJECT TIMELINE

The stakeholders are IIS as the owner and provider of the tool and the data on broadband measurements as well as ITU as the researching organisation. Contrary to other pilot countries, the Sweden country pilot did not require a mission of the data scientist. Following initial correspondence and negotiation of data access, the data was extracted by IIS and downloaded by ITU for storage and analysis.

Table 1. Project timeline.

Activity	Time
Initial communication	August 2016
Negotiation for data access	September-November 2016
Data transfer	December 2016
Data analysis	January – September 2017
Drafting of the country report	October – December 2017

2. GETTING ACCESS TO THE DATA: PROCEDURES, LEGAL DOCUMENTS AND CHALLENGES

2.1. LEGAL DOCUMENTS AND CHALLENGES

IIS is an independent public-service organization that acts to ensure positive development of the internet. The data generated from Bredbandskollen can therefore be accessed for non-commercial research purposes. However, the privacy of individuals must be protected. Any data which could disclose any sensitive individual personal data could not be published and no inquiries regarding the identity behind any individual's personal information could be asked.

Prior to the data being provided to ITU, IIS removed any personal information from the dataset, such as IP-addresses. Since detailed geographic coordinates could also be considered individual personal data, it was agreed to replace detailed geographic coordinates with LAU2 codes, which would overcome this challenge. Detailed geographic coordinates were considered not necessary for the Sweden pilot.

To access the data, a contract is generally signed between IIS as the owner of the data and the researching organization to ensure the privacy of individuals; however, because all sensitive personal information had been removed from the dataset, no contract was necessary to ensure the privacy of individuals.

2.2. RESOURCES REQUIRED FOR THE PROCESSING

The key resource is any database software which could handle large datasets of a few hundred million records could be used for the analysis. In this pilot, ITU's MSSQL database was utilized as well as Amazon Web Service (AWS) instances linked to an AWS simple cloud storage service (S3). Part of the analysis of the data was processed in a Spark cluster set in a Databricks environment.

3. METHODOLOGY, TECHNICAL DESCRIPTION, TOOLS

3.1. TECHNICAL DESCRIPTION OF MEASUREMENT METHODOLOGIES

From a technical standpoint, the most accurate option for measuring real broadband speeds is to implement in-network measurements, such as real traffic monitoring (using network counters) or test-call routines. The technical issues concerning this type of measurements are well known. Indeed, international standardization bodies, such as the European Telecommunications Standards Institute, have issued internationally agreed guidelines on how to perform this type of measurements (European Telecommunications Standards Institute, 2005). In some countries, such as India and Spain, operators are required to self-report in-network measurements and the results are disclosed to the public on a regular basis (Ministerio de Energía, Turismo y Agenda Digital, 2017; Zuhyle and Mirandilla-Santos, 2015).

However, in-network measurements can only be accomplished by those stakeholders having direct access to the network infrastructure and therefore rely on the self-reporting of the concerned network operators. Because of the difficulty of overseeing compliance and the lack of independence of this type of measurements, this approach is usually not considered a solution on its own, but rather a complement to other external measurements.

Therefore, external broadband speed measurements are the most common approach to gaining insights about real broadband speeds. There are a number of private stakeholders performing this kind of measurements and publishing data with a global coverage. These include, inter alia, Ookla's Speedtest and Akamai speed reports (Bauer et al., 2010, 2016; Lehr et al., 2013).

In addition, there exist several external broadband speed measurement platforms commissioned or operated by public entities and the academia. Some of them rely on hardware-based approaches, such as the SamKnows' Whitebox (SamKnows, 2013, 2014) or the BISmark router (Chetty et al., 2013; Sundaresan et al., 2014, 2012). However, the majority of external measurements rely on pure software approaches.

Examples of software-based approaches from public entities include the Regulatory Commission of Sri Lanka's Internet speed test platform (Zuhyle and Mirandilla-Santos, 2015), the Internet Foundation in Sweden's Bredbandskollen and the Italian Authority for Communications Guarantees' Misurainternet platform.[4]

These platforms collect additional broadband performance metrics apart from download speed, such as upload speeds and latency (delay). Depending on the type of online activity performed, some parameters will be more important than others. For instance, latency is very relevant for real-time communications, such as VoIP. As a result of the diversity of activities performed online, a complete assessment of user quality of experience cannot be summarized into a single speed metric, but will

[4] For more information, see the Misurainternet website: https://www.misurainternet.it.

require several complementary measurements (SamKnows, 2014; Wattegama and Kapugama, 2011; Zuhyle and Mirandilla-Santos, 2015).

Moreover, there exist some application-specific broadband measurements, such as Netflix's Fast.com and Youtube's Speed numbers. These measurement platforms are optimized to reflect what is the actual speed experienced by a user transferring video files with the size and protocols common in Netflix and YouTube, respectively. However, these speed measurements may be a poor approximation of the actual speed experienced when browsing a website or sending a file (Bauer et al., 2010, 2016).

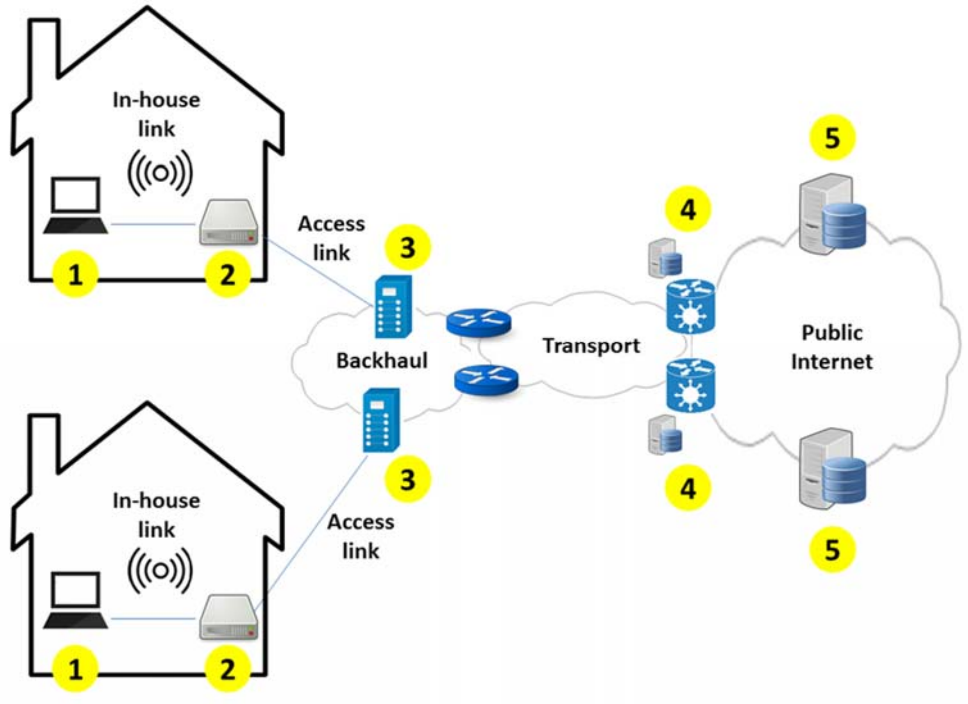
Based on a literature review of the state-of-the-art in broadband speed measurements, it can be concluded that external broadband speed measurement platforms depend on a few main design characteristics that affect their outcomes:

1. **Measurement path:** from which point to which point of the network the measurement is made (Figure 1). Speeds advertised by operators refer to the maximum achievable over the access link (points 2-3 in Figure 1). In xDSL connections, the access link is not shared, as it is for coaxial cable, but xDSL speeds are more sensitive to the distance to the local exchange (path from 2 to 3).

   From point 3 in Figure 1, there is contention for all wired technologies (fibre, coaxial, copper). That is, the transmission capacity is shared by many concurrent users and, depending on the network charge, this may lead to congestion and lower speeds. Even so, from points 2 to 4 the connection remains within the network of the ISP with whom the end user has contracted the service and therefore the speeds depend only on the network dimensioning and management of that ISP.

   End users' quality of experience is affected by the end-to-end path (i.e. points 1 to 5). This means that the speed experienced by a user also depends on the quality of the in-house connection (path from 1 to 2). For instance, the WiFi connection may not support high speeds or there might be several users making use of the same connection, therefore reducing the bandwidth available for the single user performing the test. Home network bottlenecks have been found to be very relevant in those settings in which the access is capable of providing speeds greater than 20 Mbit/s (Sundaresan et al., 2016).

Figure 1: Network diagram – key measurement points.



Source: Author based on Bauer et al. (2010).

In addition, end-to-end speeds may depend on other ISPs networks if the end user is accessing content hosted on the wider public Internet (off-net) instead of within the home ISP network (on-net). In the off-net case, the connection includes points 4 to 5 which are not under direct control of the home ISP. However, it is the decision of the home ISP with whom to interconnect and under which conditions (e.g. transit, peering). Therefore, under the assumption that customers demand universal connectivity from the ISP with whom they have the contract (see Economides, 2008, pp. 508-511), this ISP could also be hold accountable for the speed delivered over the link from 4 to 5 in Figure 1. Indeed, ISP interconnection may have a substantial impact on consumer Internet performance and business relationships between ISPs are often at the root of this performance degradation, rather than technical issues (M-Lab Research Team and others, 2014).

2. **Active versus passive testing:** passive testing measures broadband speeds over end users' normal online activities, whereas active testing relies on standardized tests run independently of each end users' actual online activity. Active testing is usually preferred in benchmarking exercises because it facilitates the comparability between measurements (Zuhyle and Mirandilla-Santos, 2015). In passive testing, if two users are performing very different activities (e.g. heavy video streaming compared with light web browsing) the outcome of the speed measurements may be significantly different even if they have the same type of connection.

3. **Voluntary versus automatic testing:** some broadband speed measurements are initiated by the end user (most software-based platforms, including Ookla and Bredbandskollen), whereas a few others are scheduled remotely and take place automatically (e.g. hardware-based platforms, such as SamKnows). Voluntary testing has several drawbacks, including the risk of selection bias. That is, end user's may run the test in a diagnostic fashion when they are experiencing network problems (Bauer et al., 2010). Moreover, actual speeds are sensitive to congestion: they tend to decrease at peak hours or during traffic burst times (Sundaresan et al., 2012; Zuhyle and Mirandilla-Santos, 2015). As a result, broadband speed results of voluntary tests may be biased because of the timing of the tests, which is out of control of the measurement platform and may not be randomly distributed. For instance, there could be more tests run during congestion periods because that is precisely when end users' perceive speed problems.

4. **Data protocol configuration:** the data transmission protocol, usually TCP, may actually be the bottleneck if not correctly configured. This is more so in high-capacity networks (e.g. gigabit broadband networks), where a single TCP flow will most likely not be enough to measure the real capacity of the link. Therefore, multiple parallel TCP flows will be required to produce a reliable measurement in high-capacity networks (Bauer et al., 2010, 2016).

5. **Statistical aggregation:** results for individual tests are calculated using different methods. These include: (i) total bytes / total time, (ii) total bytes / total time after the ramp-up period, to exclude the initial warm-up period in which transmission is slower, and (iii) payload / bytes, so that only the actual information (and not the header aggregated by the transfer protocol) is considered. In addition, different aggregation methods are used for reporting aggregate statistics. For instance, SamKnows discards the top and bottom percentiles to control for outliers and averages the rest (SamKnows, 2014). Ookla removes the fastest 10 per cent and slowest 30 per cent slices of each measurement and averages the rest (Bauer et al., 2010). Bredbandskollen takes the higher between the average speed of the whole measurement period (10s) or the average of the last 8s.[5]

6. **Other design parameters**, such as test duration time, technology used in the application (e.g. HTML or Flash-based) and the file size used in download/upload measurements (Bauer et al., 2010; Zuhyle and Mirandilla-Santos, 2015).

Table 2 summarizes how these design parameters vary according to the broadband speed measurement platforms chosen in this project to benchmark the results obtained with Bredbandskollen.

The main differences would be the following: SamKnows ability to exclude the home network from the measurement and to automatically control the test timing, which will lead to more robust results; Akamai's passive testing, which may lead to comparing apples and oranges if their results are used to

[5] See Bredbandskollen methodological FAQ, available at: https://ensupport.bredbandskollen.se/support/solutions/articles/1000228167-how-does-the-measurement-work-in-technical-terms-

benchmark the outcome of active tests; the different statistical aggregation procedures which may potentially have a significant impact on the results.

Another relevant factor which needs to be considered is the different sample sizes of each broadband platform. For instance, concerning data from Sweden (i.e. the object of this project), Bredbandskollen collected some 15 million observations per year for download speeds, Akamai 20 million and SamKnows 0.52 million. These differences illustrate the trade-off between volumes of data and measurement precision. Indeed, hardware-based measurements are costly and this severely limits the sample size even in developed countries.

Table 2: Main differences between selected Internet measurement platforms.

	Measurement path	Active/ passive	Voluntary/ automatic	Data flows	Statistical aggregation
Akamai	1-4 or 1-5 depending on server load	Passive	Real traffic	Sequential	Unknown
Bredbandskollen	1-4 in most cases. 1-5 for the rest	Active	Voluntary	Parallel	Avg.first 2s or avg 10s
Ookla	1-4 or 1-5 depending on user location	Active	Voluntary	Parallel	Avg. after ramp-up. Excl. top 10% and bottom 30% slices
SamKnows	2-5	Active	Automatic	Parallel	Avg. after ramp-up. Excl. top and bottom percentiles. Separate peak / off-peak

Source: Author based on Bauer et al. (2010, 2016); Canadi et al. (2012); SamKnows (2014) and Bredbandskollen.

3.2. STATISTICAL ISSUES

Broadband speed measurements are composed of non-representative Internet data and, as such, face similar problems to those found in non-probabilistic samples (Couper, 2013; Zagheni and Weber, 2015).

1. **Quota-based samples:** a panel is selected trying to adjust its composition according to the social, demographic and/or geographic characteristics of the in-scope population. More in general, quota-sampling is based on the selection of the surveyed individuals based on a series of covariates that are believed to explain the target variable. For instance, in its quota-based sampling methodology, SamKnows considers country, ISP and connection technology as the covariates used to define the quotas (SamKnows, 2014). Quota sampling in itself does not solve the issue of selection bias, particularly when participation is voluntary. This is indeed the case in most broadband speed measurement platforms, including SamKnows.

2. **Online voluntary surveys:** on top of the selection bias, unequal participation of the volunteers may add more uncertainty about the representativeness of the results. Indeed, there is evidence from the survey research domain pointing that a relative large numbers of online surveys are completed by a relative small number of active panelists (Couper, 2013). In the domain of broadband speed measurements, in some platforms where it is possible to identify individual end users, such as in Bredbandskollen, there is evidence that some active users may perform thousands of tests per year, whereas some others only a few. In some other broadband speed measurement platforms, however, one-time test users were found to be predominant (Wattegama and Kapugama, 2011).

From all statistical issues, selection bias is the central one when trying to make statistical inferences from the data produced by a broadband speed measurement platform. For instance, Rood et al. (2012) used official statistics on geography, technology, gender and age to test whether self-selected measurements of a broadband speed measurement platform (iPing) were representative of the whole population.[6] The results rejected this hypothesis, and the authors concluded that higher speed connections, males and persons aged 30-60 were overrepresented. Moreover, the paper found that the results of Ookla and Samknows in the Netherlands also risked not being representative of the whole population. Several approaches have been proposed to deal with selection bias:

1. **Selection panel:** create a proper selection panel from which to draw detailed socioeconomic household and individual data in addition to the broadband measurements. This selection panel can then be used to weight the results of the self-selected panel and produce representative results out of it (Rood et al., 2012). This is indeed a reliable but costly option, because it requires the maintenance of a regular panel in addition to the upkeep of the speed measurement application.

2. **Calibration against ground-truth data:** use data from a reliable source (i.e. the so called ground-truth data) to weight the results of the broadband speed test in order to ensure that the results of the two sources correspond for a period in which they overlap. This correction relies on the assumption that there is a (potentially stochastic) structural relationship between the subgroup of the population included in the broadband speed test sample and those not included. Moreover, this relationship needs to hold in time (Zagheni and Weber, 2015). Couper (2013) warns that there is a risk of in-sample overfitting and type I errors when calibrating large datasets.

3. **Differences-in-differences approach:** when it is not possible to obtain a robust point in time estimation because the previous approaches do not apply, it may still be possible to obtain a sound indication of trends by applying a differences-in-differences approach (Zagheni and Weber, 2015). For such an approach to be feasible, a control group needs to be identified and

[6] Official statistics on geography, technology, gender and age in the Netherlands are produced via household surveys by Statistics Netherlands (CBS) and administrative regulatory filings of ISPs to the Dutch regulator OPTA.

the selection bias in the control and the measurement group needs to be constant with time or evolve in parallel.[7] In addition, analogous to calibration, a key assumption needed to validate this approach is that relative changes for the subgroups of the population included in the sample should be indicative of trends in the general population.

Based on these points, a key question that needs to be addressed when analyzing broadband speed measurements is how volunteers taking part in the test differ from non-volunteers. This is a common question to most non-probabilistic survey research (Couper, 2013).

In practical terms, the statistical issues highlighted in this section often cannot be addressed *a posteriori*; that is, during the data processing stage where the available information is fixed.

For instance, a common problem when analyzing software-based broadband speed tests is that the information on the number of unique users is not available or not disclosed to the researchers (Canadi et al., 2012). This is also the case in the Bredbandskollen microdata that IIS shared with ITU, which for privacy issues does not contain a unique identifier for each user, but rather for each test. The lack of this information makes it difficult to perform any ex-post calibration, because the cross-linking of the observations from the speed measurement platform with external individual population characteristics is not possible.

When facing major problems inherent to the structure of the underlying dataset, researchers often follow the approach of describing the limitations of the dataset and publishing the results as such, with little (Chetty et al., 2013; Prasad et al., 2016; Wattegama and Kapugama, 2011) or some admonition about their representativeness for the whole population (Canadi et al., 2012; Riddlesden and Singleton, 2014). In the latter, an approximate correspondence between the global average speed obtained in the study and that obtained in some previous more robust measurement effort is considered as a broad indication of the validity of the results.

More generally, the validity of the statistical inference derived from each broadband speed measurement platform may depend on the purpose of the research. If the objective is to raise awareness that some consumers are not getting the speeds specified in their contracts, showing that this is true for a large enough number of end users' may suffice (Chetty et al., 2013; Wattegama and Kapugama, 2011).

If the purpose of a broadband speed measurement platform is customer self-check, a sound measurement methodology (as discussed in Section 2.1) and consistency in the results for a single user is all that is required. This is, for instance, the primary objective of Ookla's Speedtest. The Italian Authority for Communications Guarantees' Misurainternet platform incorporates this logic in the application itself. Indeed, by means of an electronic certificate, legally validated results of the speed test

[7] For example, there may be a selection bias in the control and measurement groups towards an over-representation of end users from a given operator. If at some point in the time series the selection bias changes only in the control group (e.g. there is no longer an over-representation of clients from that operator), then the differences-in-differences approach would no longer be valid.

can be communicated to the concerned ISP who then has the obligation to reestablish a minimum speed according to the conditions specified in the contract.

If the objective of the research is to obtain representative information on the real broadband speeds experienced by broadband subscribers in a country, as it is the case in official telecommunication statistics, then all the statistical considerations discussed in this section should be addressed.

Lastly, broadband speed tests can also be a rich source of data for analyzing different Internet conditions and user behaviors across socio-economic strata. An obstacle to mapping broadband speed data to external data sources is the fact that broadband tests usually do not collect demographic information. Indeed, socio-demographic information is key to linking technical measurements to wider societal issues (for an example, see Chapter 6 in International Telecommunication Union, 2016).

Some broadband speed measurement platforms have bypassed this problem by asking end users to provide extra voluntary information, such as their postal code. Based on this extra information, it is possible to geolocate each test and link it with indicators of socio-spatial structures, such as rurality and levels of material deprivation (Riddlesden and Singleton, 2014). Other platforms ask for demographic data, such as age and gender, on a voluntary basis after or prior to the test. This information allows the linkage of the test results with demographic information from household surveys (Rood et al., 2012).

3.3. DESCRIPTION OF THE DATA USED IN THE PROJECT

ITU collects data on fixed-broadband subscriptions broken down by advertised speed. Advertised speeds, however, provide only an upper-limit to the actual broadband speeds. Advertised speeds are determined by each service provider according to their own internal methodology, which is not disclosed nor usually inspected by an independent party. However, it is understood that these speeds indicate the maximum or peak data rates that a customer may experience in the link between the customer location and the broadband provider (Bauer et al., 2010).

On the other hand, precise external hardware-based measurements, such as the ones commissioned to SamKnows by the regulatory agencies in the UK (Ofcom, 2017), the United States (Federal Communications Commission, 2015b) and the European Commission (SamKnows, 2014) are costly. Therefore, they cannot be realistically scaled up to a wider set of countries. Software-based, crowdsourcing data on Internet speed measurements remains the only possible stable source of real broadband speed information for most countries. Moreover, the low cost of deployment of these measurement platforms makes it possible to envisage its adoption by any interested regulator/policy-making.

Some regulators in developing countries, such as the Telecommunications Regulatory Commission of Sri Lanka, have already launched their own measurement portal. There are also private stakeholders, such as Ookla, recording these data at the global level. In this context, the Internet Foundation in Sweden has been a forerunner of public-service broadband measurement platforms with its portal Bredbandskollen (Davidsson, 2017). Bredbandskollen is a free online software tool. Internet users can measure their real broadband speeds for both fixed and mobile connections by running the test either on its website or by downloading an app and executing the test from a mobile device. Broadband speed tests are initiated by the end user, whenever they click on the start test button on the website or launch the measurement from the mobile app.

Every day, about 100,000 measurements are conducted, mainly in Sweden, but data also exists from other countries. In total, over 200 million measurements have been made since 2007. This project takes advantage of the microdata made available by IIS to ITU on the broadband speed measurements from the Bredbandskollen platform in the last six years (2011-2016). The measurements are not representative of the population as some active users may perform thousands of tests per year, whereas some others only a few. Information on the number of unique users is not available because of privacy issues. The variables contains in the dataset are shown in Table 3. Figure 2 shows a sample of the data used to compute the indicators.

Table 3: Data record for each observation

Field	Description
Unique ID	Unique number identifying each test. A single user running several tests will be assigned several unique IDs, i.e. one for each test.
Client	iOS (mobile app for iOS), Android (mobile app for Android) or Web measurement from the website).
Date/time	UTC time, i.e. 1 or 2 hours less than Swedish time depending on the period of the year.
Download speed	Result of the test in Mbit/s.
Upload speed	Result of the test in Mbit/s.
Avg. response time	Result of the latency test in milliseconds.
Country code	Country code of the location where the test was conducted. Empty if the location of the measurement is not available.
Region code	Region code of the location of the measurement. Blank if outside Sweden.
Municipality code	Municipality code of the location of the measurement. Blank if outside Sweden.
Municipality	Name of the municipality. May be empty.
Network type	May be empty. If the client type is iOS or Android, the following options are possible: CDMA, GPRS, EDGE, UMTS, HSPA, HSDPA, HSPAP, LTE, Mobile (if the network type is not detected) and WiFi (measurement via WiFi or otherwise not using the mobile network). If the client type is Web, there is some description of the subscription. Often it is based on information provided by the user, so it is not always reliable.
Operator	Operator name, inferred from the IP number. Usually correct, but not always reliable. May be empty.
Phone model	Description of the phone model, if the client type is iOS or Android. If the client type is Web, description of the web browsers.

Figure 2: Sample of the data

UniqueID	Client	DateTime	Download Speed	Upload Speed	AvResponse Time	Country code	Region code	Kommunc ode	Kommun	NetworkT ype	Operator	PhoneModel
48770254	iOS	15/07/2011 22:36	0.2	1.8	380	TR	0	0		WIFI	Turk Telekom	iPhone3,1
210601774	iOS	18/10/2016 08:03	30.6	11.4	37	SE	10	1082	Karlshamn	LTE	Hi3G Access AB	iPhone7,2
47747325	Android	25/06/2011 12:05	1.0	1.9	112	SE	14	1419		MOBILE	TeliaSonera AB	X10i
51681581	Web	06/09/2011 21:10	92.6	11.4	0	SE	23	2380	Ã–stersund	60-100 Mbit/s fiber	Bahnhof Internet AB	Mozilla/5.0 (Windows NT 5.1) AppleWebKit/535.1 (KHTML, like Gecko) Chrome/13.0.782.220 Safari/535.1
48664519	iOS	14/07/2011 06:32	0.5	1.1	373		0	0		WIFI	Hi3G Access AB	iPad2,1
210665998	iOS	18/10/2016 18:37	62.4	12.0	12		0	0		WIFI	Telenor Sverige AB	iPhone8,1
210666012	Web	18/10/2016 18:37	32.4	9.5	31	SE	22	2281	Sundsvall	150-250 Mbit/s kabeltv	Com Hem AB	Mozilla/5.0 (Macintosh, Intel Mac OS X 10_8_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.112 Safari/537.36
210666000	Web	18/10/2016 18:37	93.3	95.2	7	SE	20	2080	Falun	500-1000 Mbit/s fiber	Net IT Internet Solutions in Sweden AB	Mozilla/5.0 (Windows NT 10.0, WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/53.0.2785.143 Safari/537.36
51681571	Web	06/09/2011 21:10	6.0	1.6	25	SE	1	120		12-24 Mbit/s adsl	TeliaSonera AB	Mozilla/5.0 (Windows NT 6.1, WOW64) AppleWebKit/535.1 (KHTML, like Gecko) Chrome/13.0.782.220 Safari/535.1
134434826	Web	21/05/2014 16:12	5.7	0.7	30	SE	1	181	SÃ¶dertÃ¤lje		Bredbandsbolaget AB	Mozilla/5.0 (Windows NT 6.3, WOW64Â§ rv:29.0) Gecko/20100101 Firefox/29.0
134434847	iOS	21/05/2014 16:13	1.4	1.2	92	SE	3	382		MOBILE	Tele2 Mobile	iPad2,2

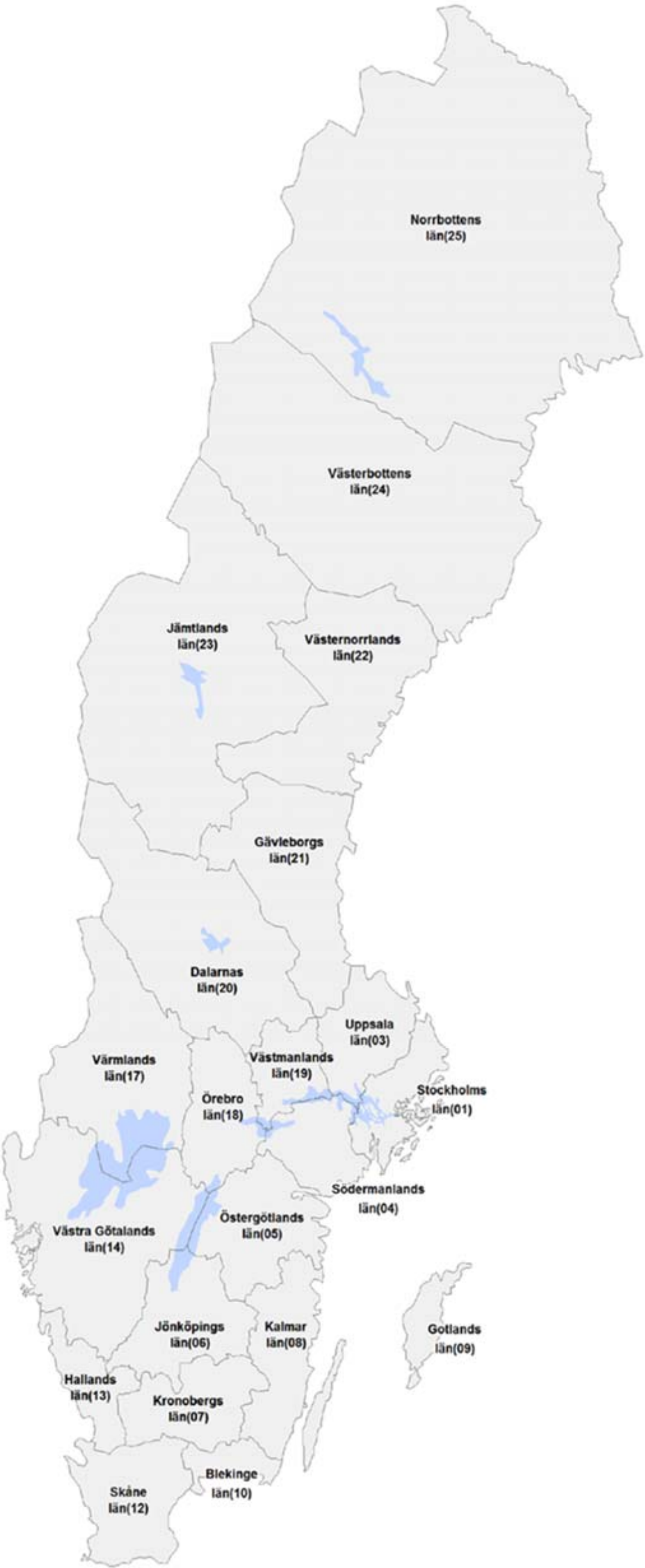
3.4. REFERENCE DATA USED IN THE PROJECT

Reference data are required to be able to fully conduct the project. The reference data are used to link the data to geographical location and to calculate some specific indicators. The administrative borders of Sweden's subdivisions (Local Administrative Units (LAUs)) were provided at two levels. In Sweden there are 21 administrative regions at LAU1 level and 290 municipalities at LAU2 level. Figure 3 presents the LAU1 regions in Sweden.

Reference data also included official indicators on telecommunications/ICTs collected by ITU. ITU's data collection is divided into two categories according to the source of the data collection: administrative data, and household surveys or censuses. The methodologies for collecting harmonized and internationally comparable ICT data from these two data sources are available on ITU's website.

The measurements from Bredbandskollen were also compared with those obtained by other broadband speed measurement platforms. Data on broadband speed measurements for Sweden were therefore also obtained from Akamai, Ookla and SamKnows.

Figure 3: LAU1 regions of Sweden



3.5. DATA ACCESSING AND PROCESSING

Given the large size of the dataset (more than 100 million observations corresponding to about 30 GB of data), the data processing was performed in a Spark cluster set in a Databricks environment.[8, 9] The cluster was composed of one driver and up to eight worker nodes.[10] The worker nodes were auto-scaled according to the computing requirements of each job submitted.

The records were extracted into comma-separated values ("csv") format from the SQL database. These files were downloaded directly into an Amazon Web Service (AWS) instance and transfered to an AWS simple cloud storage service (S3) using multipart upload.[11] The S3 data bucket was mounted on the Databricks environment, thus linking the Spark cluster with the raw ".csv" files. The heavy data processing was carried out in the Spark cluster and the summary results saved back into ".csv" files. These files were transfered to a local computer and used to produce the charts using R. Figure 4 shows a graphical summary of the data processing environments used in this project and the corresponding data flows.

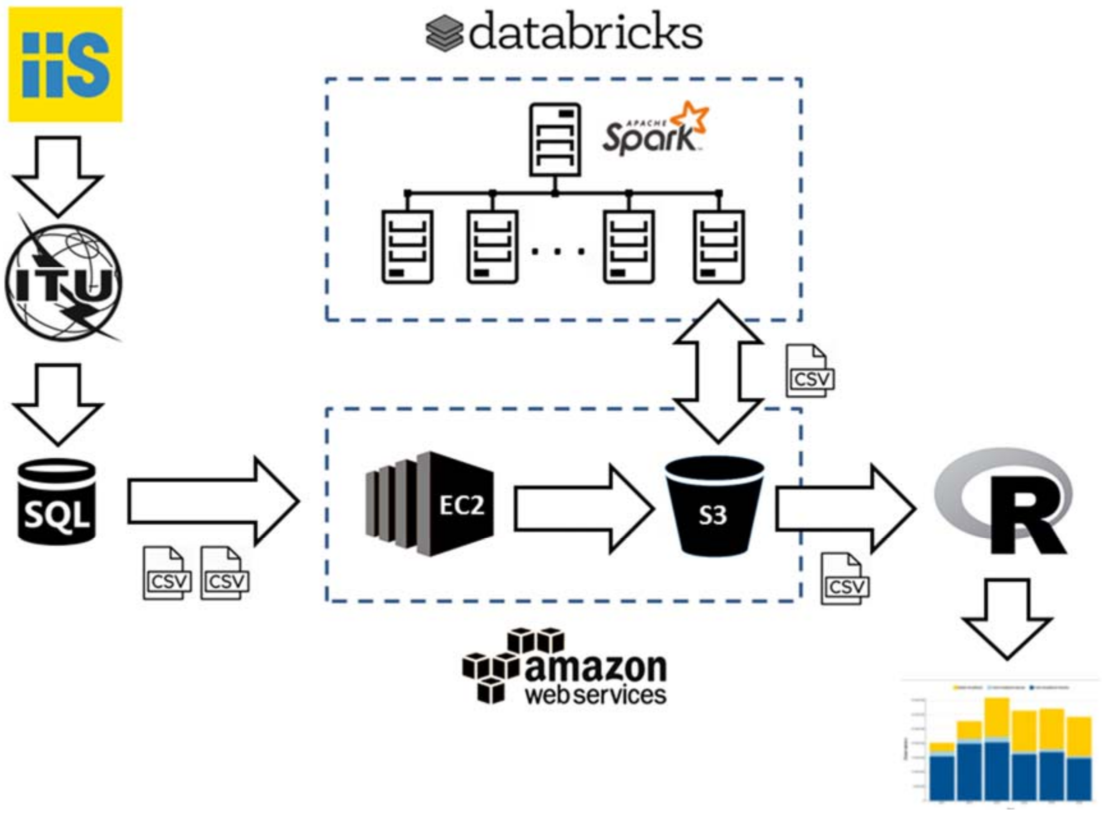
[8] Apache Spark is an open-source cluster computing framework that provides an interface for largescale data processing with implicit parallelism. In comparison to other cluster-computing platforms, such as Hadoop Map Reduce, Spark is significantly faster because, inter alia, it exploits in memory computing. Moreover, Spark provides a well maintained Python API (PySpark) which makes it possible to program Spark clusters using Python-like syntax.

[9] Databricks is an online analytics platform that provides environments for cloud-based big data processing using Apache Spark. For more information, see Databricks website: https://databricks.com.

[10] Each node was an Amazon Web Service instance of the type r3.xlarge, with 30.5 GB of memory, four cores and one DBU.

[11] AWS multipart upload enables to upload large files in parts. For more information, see the AWS webpage on this topic: http://docs.aws.amazon.com/AmazonS3/latest/dev/mpuoverview.html.

Figure 4: Data processing environment



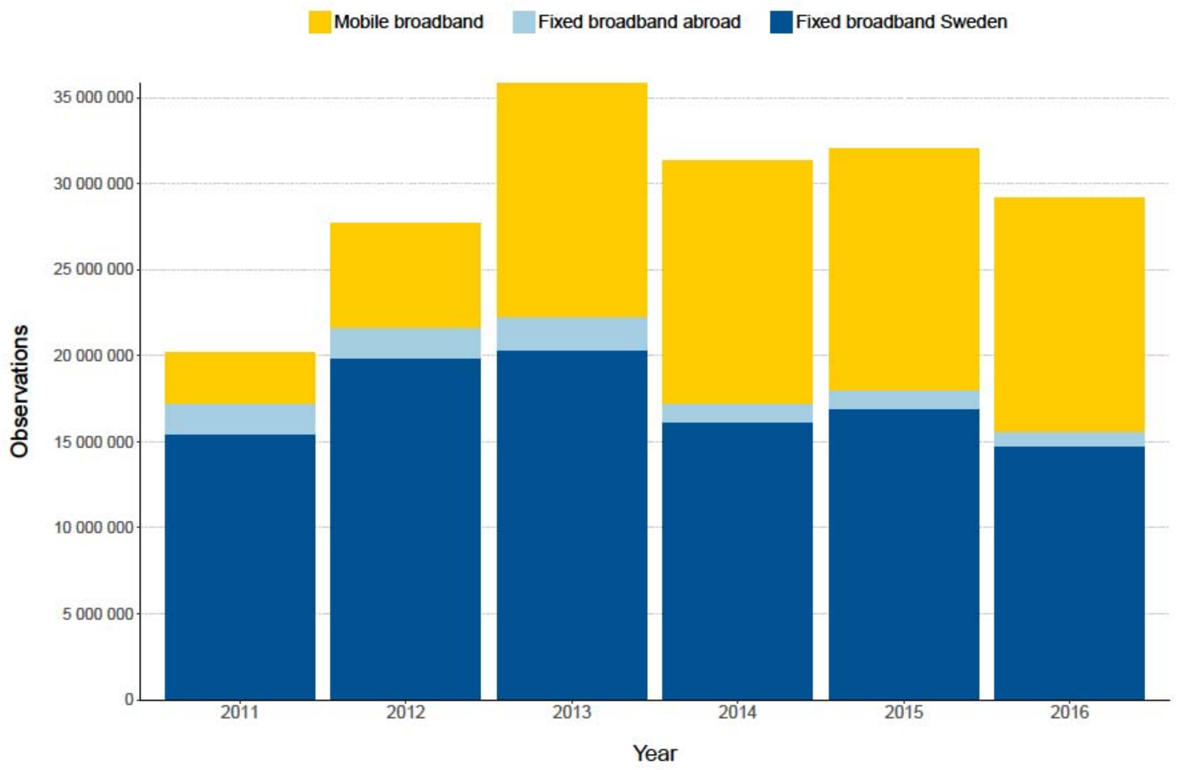
4. RESULTS DERIVED FOR THE BIG DATA INDICATORS

The dataset used for the analysis in this project includes around 15 million test observations per year for fixed-broadband connections (Figure 5). Mobile-broadband observations were not considered, because they are affected by different technical determinants and would require a different, more challenging processing.

Observations from fixed-broadband tests performed outside the country are also excluded, given that they concern the quality of the access network of another country and therefore require a separate analysis. In 2016, more than half of the speed tests outside of Sweden were made in other Nordic countries.

The fixed-broadband speed tests ran in Sweden are filtered to exclude those in which download speed, upload speed or latency where not greater than zero. These are considered as erroneous measurements and included in the light blue bar in Figure 5.

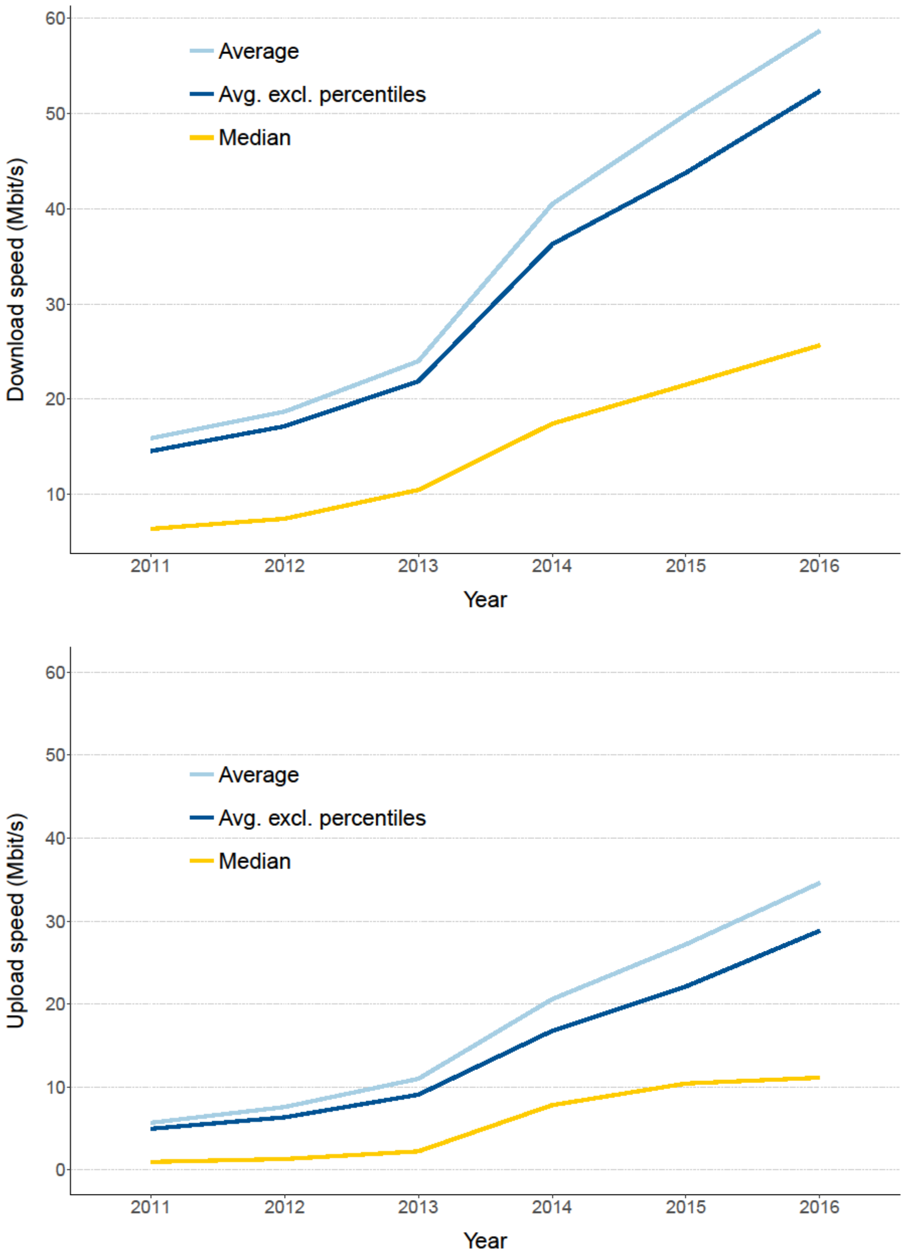
Figure 5: Number of observations, by year and type of service.



4.1. OVERALL ANALYSIS

Figure 6 shows the averages for download and upload speeds of all valid test results performed from a fixed-broadband connection in Sweden. In addition to the average, the median is presented as well as the result of removing the top and bottom percentiles and averaging the rest. This latter approach is proposed by SamKnows (2014) in order screen anomalous or misrepresentative test results.

Figure 6: Fixed-broadband download speeds (top) and upload speeds (bottom), 2011-2016



Note: Avg. excl. percentiles refers to the average obtained after removing the top and bottom percentiles

There is a significant difference between the median and the mean. The number of observations with speeds above 100 Mbit/s drags up the average while many observations still have speed well below 10 Mbit/s. This trend is increasing over time as access to higher speeds for some is rising faster than the speed most people can access.

The histograms also show the evolution in the distribution of observed speeds between 2011 and 2016. For instance, download speeds above 30 Mbit/s – the threshold considered by the European Commission for high speeds in the Digital agenda (European Commission, 2010) – represented around 10 per cent of all observations in 2011, whereas in 2016 they surpassed 40 per cent. As could be expected, upload speeds tend to be lower than download speeds: only 15 per cent of all observations had upload speeds above 30 Mbit/s in 2016, and in 2011 most observations had upload speeds in the interval 0-2 Mbit/s.

Figure 7: Histograms of fixed-broadband download speeds (top) and upload speeds (bottom), Sweden – Bredbandskollen test results.

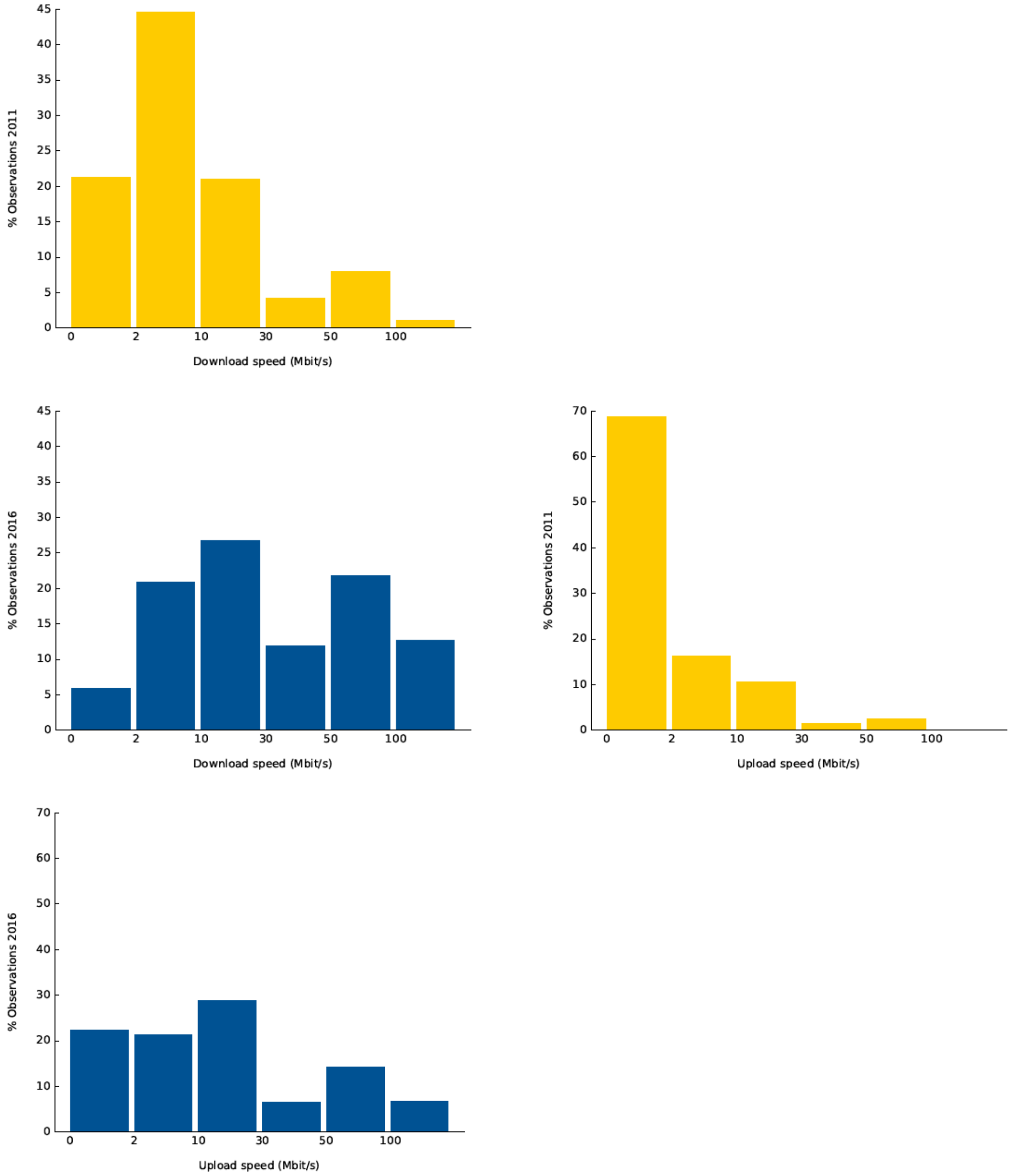
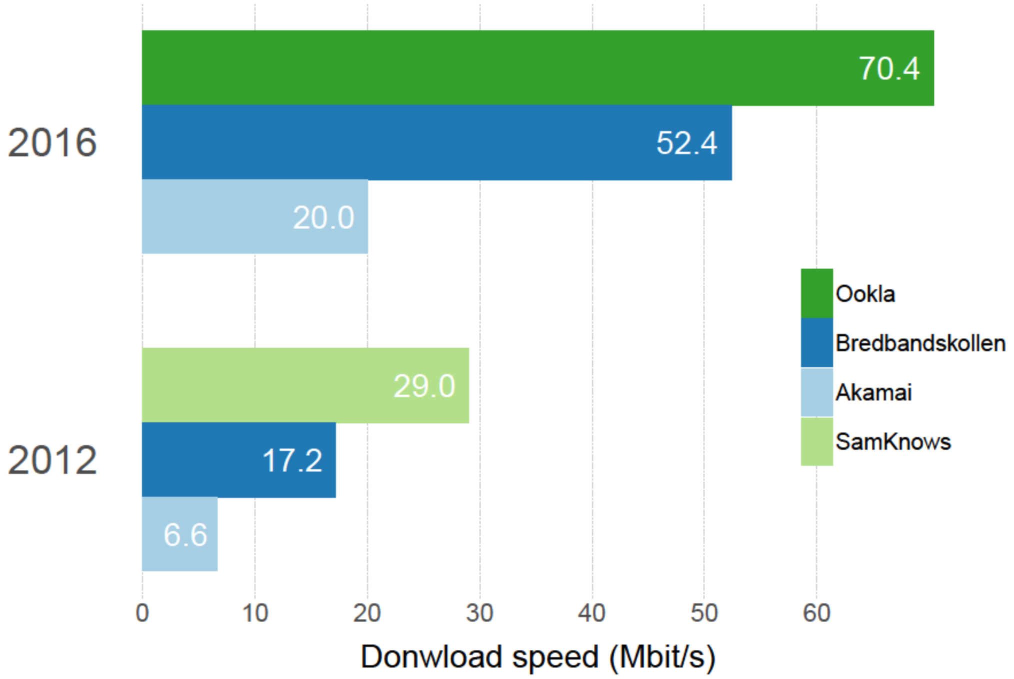


Figure 8 compares the results of Bredbandskollen with those obtained by other broadband speed measurement platforms. For Bredbandskollen, the average excluding the top and bottom percentile is used hereafter in this section. This aggregation method is preferred because it is the same as applied by (SamKnows, 2014). Some filtering is required to screen anomalous results and, indeed, most broadband speed measurement platforms apply some sort of filtering (see Table 2).

Figure 8: Fixed-broadband download speeds, Sweden, 2012 and 2016 – comparison of Bredbandskollen results against other measurement platforms.



There are large differences between measurement platforms: Ookla's and Samknow's averages are, respectively, 34 per cent and 69 per cent higher than Bredbandskollen's. Akamai's average is about 62 per cent cent lower than Bredbandskollen's.

For the only two measurement platforms for which we have several years of overlapping results, Akamai and Bredbandskollen, the relative difference between the two averages remains rather constant with time. Indeed, Akamai speed averages are in the range of 56 to 64 per cent lower than Bredbandskollen in the period 2011-2016. Both platforms have large and comparable sample sizes (above 15 million observations per year), and therefore the difference observed is probably due to structural persisting disparities in the measurement methodology. For instance, Akamai's passive monitoring of speeds as opposed to Bredbandskollen's user-initiated active testing.

The average speed measured by Bredbandskollen is also significantly lower than that obtained by SamKnows in 2012. SamKnows is often taken as ground-truth data in studies targeting countries with large SamKnows deployments, such as the United States or the United Kingdom, because of the robustness of their hardware-based methodology (Canadi et al., 2012; Riddlesden and Singleton, 2014).

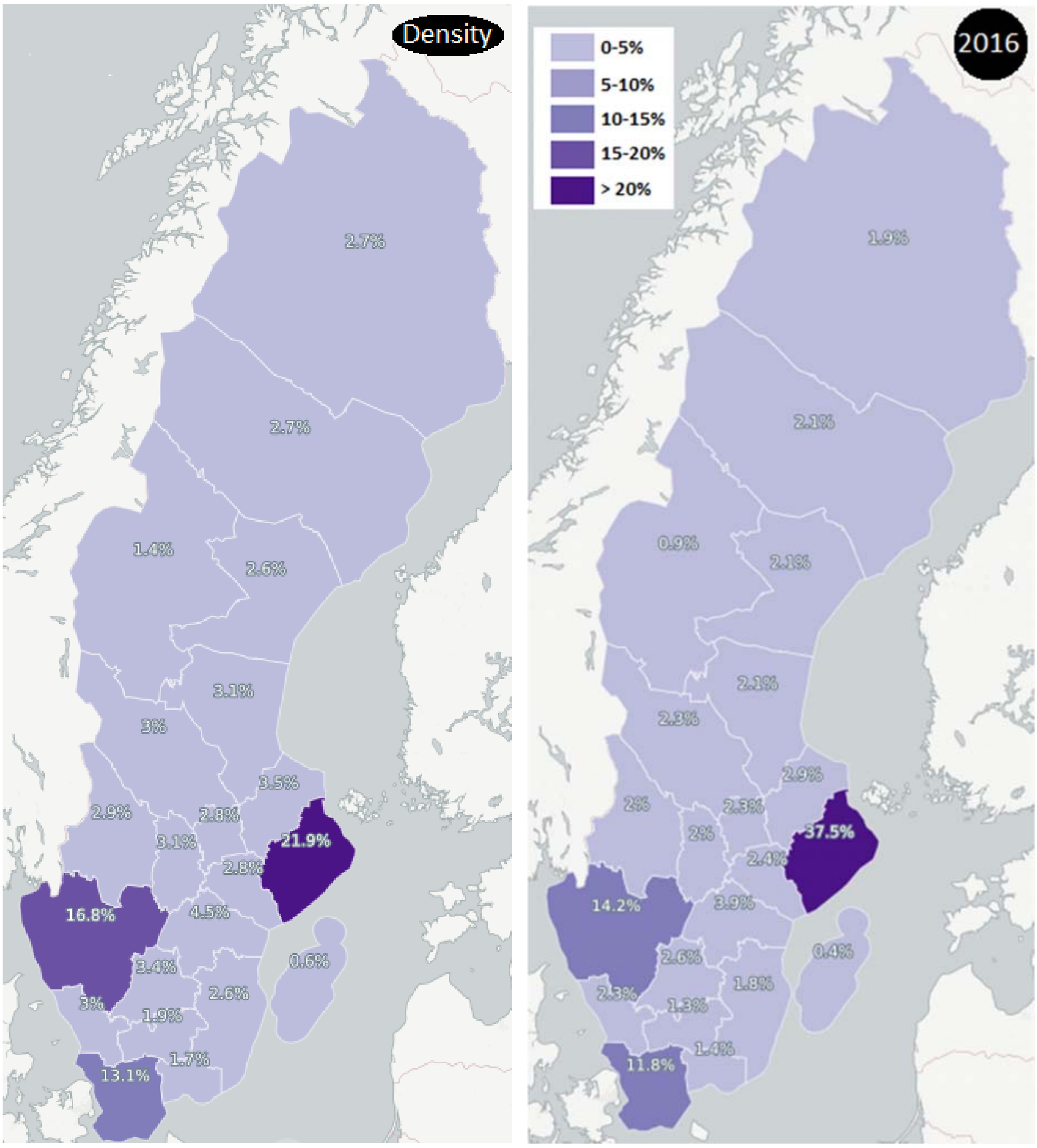
In the case of Sweden, it is not warranted that SamKnows can be used as ground-truth data because of its relative small sample size in Sweden (about 0.5 million observations per year coming from 217 end users). Indeed, in the studies commissioned to SamKnows by the European Union, only partial results are published for Sweden (SamKnows, 2013, 2014).[12] Nevertheless, SamKnows remains an important reference given that it is the only benchmark for Sweden in which possible bottlenecks within the home network (e.g. concurrent use of a single connection, poor WiFi performance) are discounted. SamKnows average speed is well above that measured by Bredbandskollen, which may indicate that congestion within the home network (path from 1 to 2 in Figure 2) is a relevant bottleneck in Sweden.

4.2. GEOGRAPHIC ANALYSIS

As expected, most observations are concentrated around the most populous areas of Sweden with more than one-third of all observations in the Stockholm region (Figure 9). Together with Västra Götalands region and Skåne region, with the second and third largest cities of Gothenburg and Malmo, the top population centers of Sweden represents nearly two-thirds of all observations. This is in line with the population densities of these three regions, which concentrate most of the households with access to fixed-broadband services in Sweden. However, Stockholm region seems to be over-represented, because it accumulated 37.5 per cent of all observations in 2016 whereas it only represents 21.9 per cent of the households with potential access to a fixed-broadband connection.

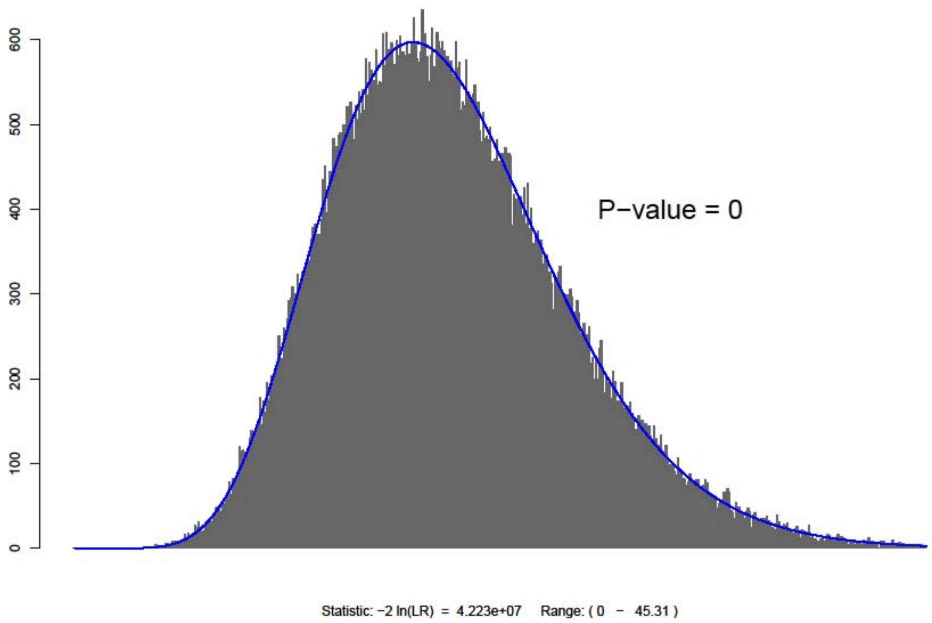
[12] For 2012, SamKnows published the average download speeds for FTTx and xDSL connections in Sweden, but not for cable. For 2013, only the average download speed of FTTx connections was published. In Figure 9, the average for SamKnows is estimated by taking for cable connections in Sweden the average speed of cable connections in Europe in 2012. This is a conservative estimate, given that average speeds in Sweden tend to be above the European average.

Figure 9: Household density (left) and share of Bredbandskollen observations (right) per region, 2016.



In order to know precisely whether its representation in the sample is correct or not, data on the number of fixed-broadband subscriptions would be needed to be broken down by region, which unfortunately PTS does not collect. However, a chi-squared test to check whether the sample distribution per region corresponds to the household density can be applied. The results strongly reject the null hypothesis that both come from the same distribution (Figure 10).

Figure 10: Chi-squared test: share of Bredbandskollen observations and household density per region.



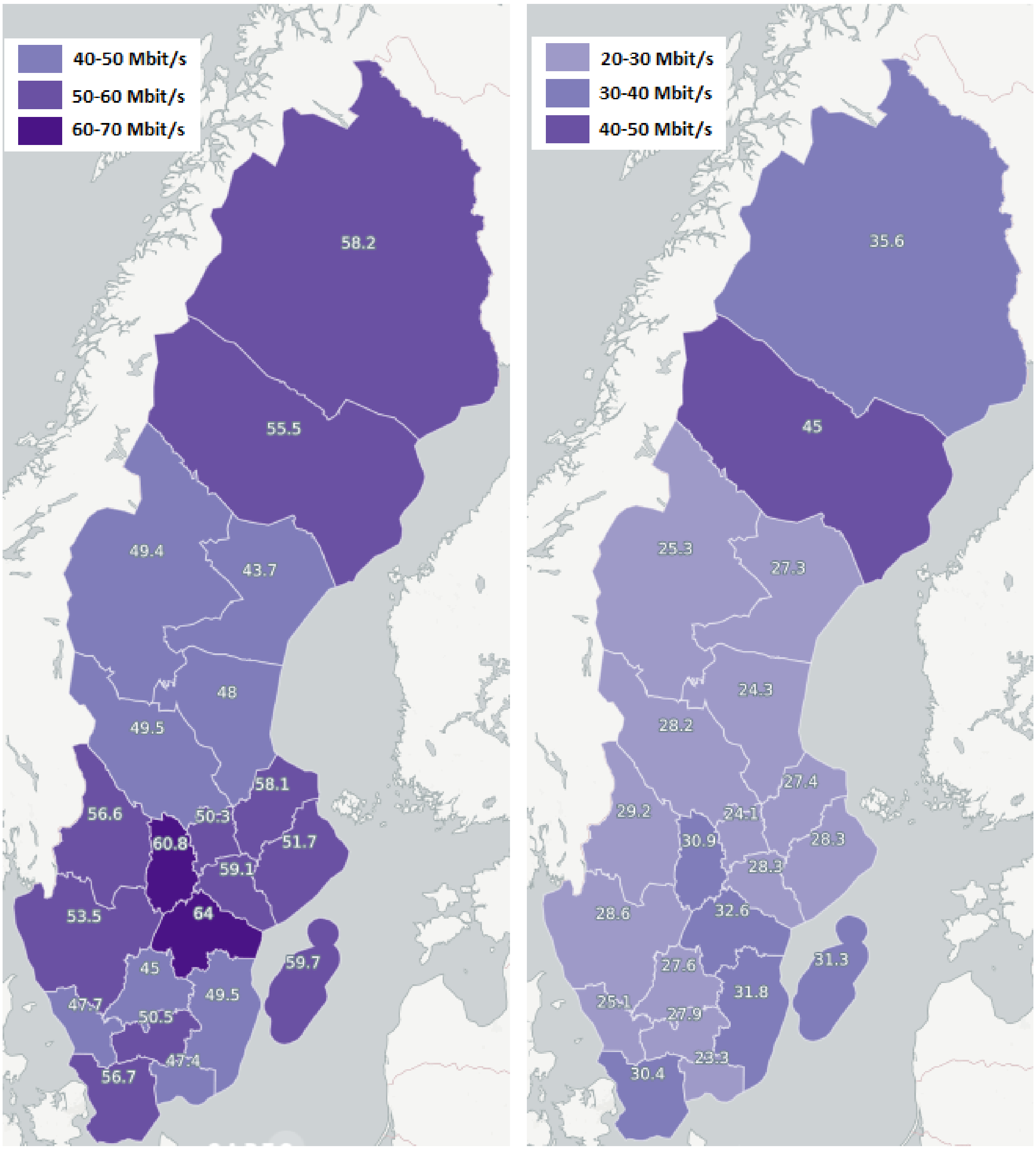
It can be assumed that the share of subscriptions per region corresponds to the share of potential customers per region. This is often not the case, as dense urban areas tend to concentrate most of the connections, whereas sparse rural areas are less connected, even proportionally to their population (see chapter 6 in International Telecommunication Union, 2016).

A striking finding from the geographic analysis is that the regions with the highest population densities (i.e. Stockholm, Västra Götalands and Skåne) are not the ones with the highest average speeds. On the contrary, the highest average upload speeds (and also some of the highest download speeds) are recorded in Norrbotten and Västerbotten, two sparsely inhabited regions at the uppermost part of Sweden. The cold conditions and cheap and clean energy in the north of Sweden present a hospitable environment for data servers, and over the past decade, these regions have been successful in attracting large data centers. Figure 11 shows the average broadband speeds recorded by Bredbandskollen in each region in 2016.

There is nothing in the analysis of Bredbandskollen's observations indicating an anomaly in the representation of these regions in the sample. Therefore, this surprising result seems to be correct and would need to be cross-validated against some qualitative evidence.

Another message that emerges from the map of average speeds in Sweden is the relative equality of speeds in all regions. Moreover, this equality is grounded on a very high baseline: 40 Mbit/s download and 20 Mbit/s upload, as the minimum averages per region. This is uncommon and suggests that Sweden has had much more success than most other countries in ensuring that the benefits of high-speed broadband reach rural and sparsely populated areas.

Figure 11: Fixed-broadband download speeds (left) and upload speeds (right), by region, Sweden, 2016 – Bredbandskollen test results.



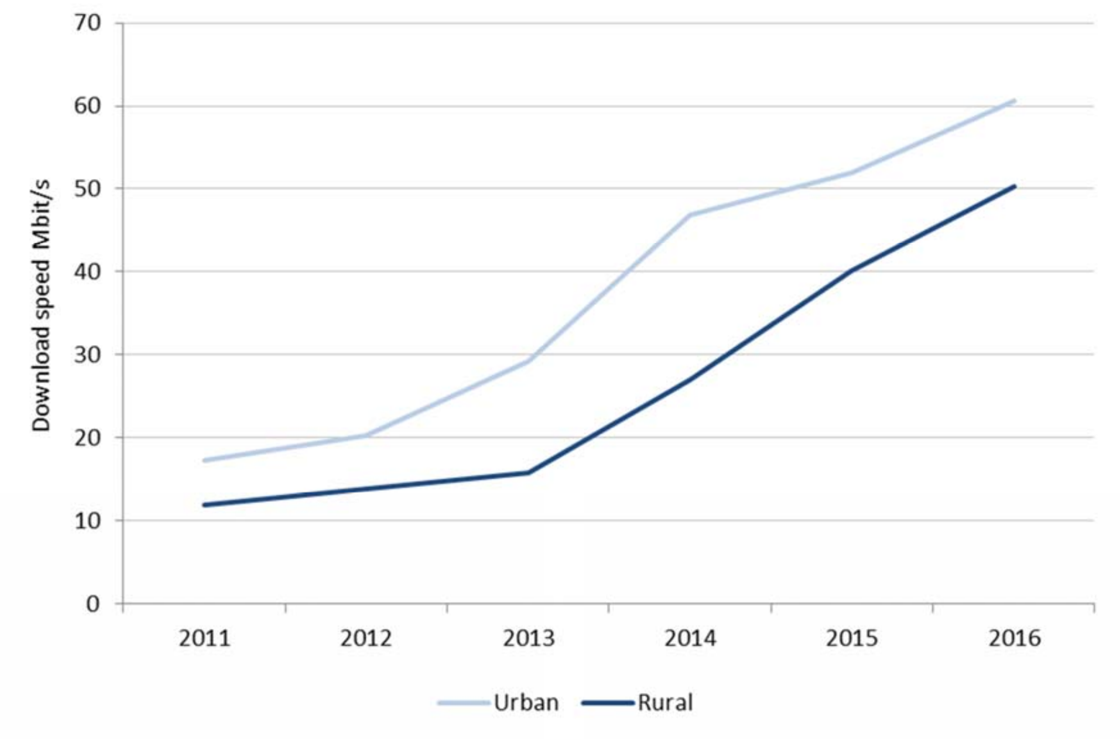
There are several definitions in Sweden of how to classify urban and rural areas. Statistics Sweden defines an urban area as a locality (group of houses) with at least 200 inhabitants. According to this definition, around 85 per cent of the Swedish population lives in an urban area. The OECD definition defines only areas with a population density greater than 150 inhabitants per square kilometre as urban

(OECD, 2011). According to this definition, only 30 per cent of the Swedish population lives in an urban area.

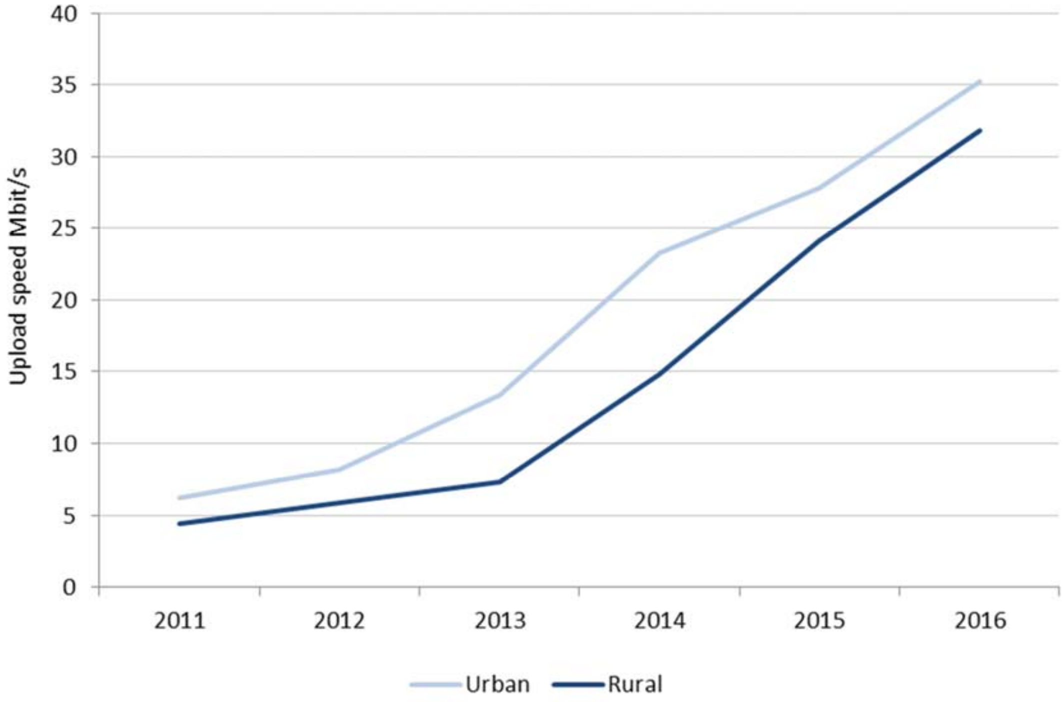
Since the data from Bredbandskollen is only available at LAU2 level, none of these definitions could be used. Instead, this report uses a definition from Swedish Association of Local Authorities and Regions, which classifies the municipalities in 9 different categories.[13] These categories are (1) Big Cities (2) Big city suburbs (3) Larger city (4) Commuting municipality near major city (5) Low-cost municipality near major city (6) Small town / town (7) Commuting municipality near small town / town (8) villages and (9) Rural municipality with visiting industry.

For this report, the categories 1, 2, 3 and 6 (cities, suburbs and towns) are classified as urban areas, while commuting municipalities and villages are classified as rural areas. With this definition, around 72% of the Swedish population lives in an urban area and 28% in rural areas. Figure 12 shows the differences in average download and upload speed between urban and rural areas from 2011 to 2016.

Figure 12: Fixed-broadband download speeds (top) and upload speeds (bottom) by urban and rural, 2011-2016



[13] For more information about this classification, see https://skl.se/tjanster/kommunerlandsting/faktakommunerochlandsting/kommungruppsindelning.2051.html (in Swedish).

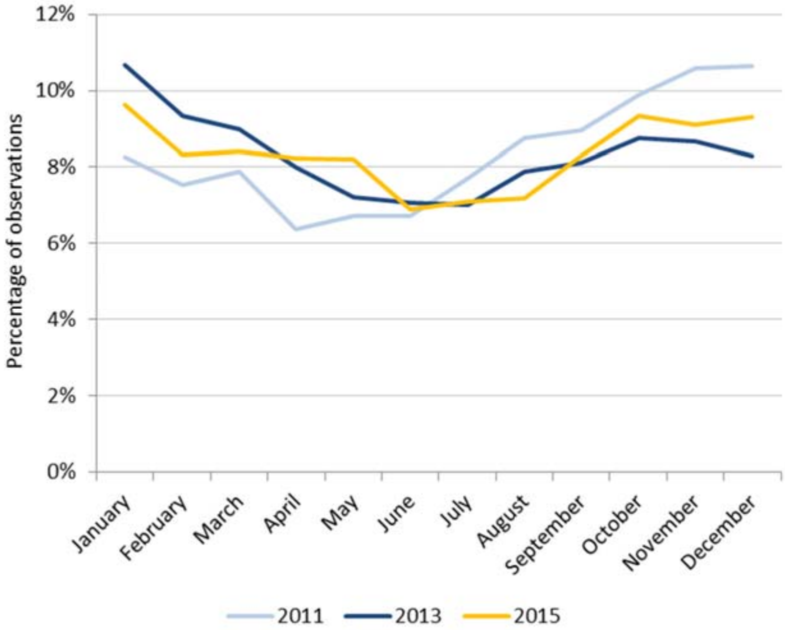


4.3. TIME ANALYSIS

In order to conduct an analysis of speed at different times of the day, the time variable, which provided the time of observation in UTC time, has been converted to actual Swedish time. This conversion also included taking into account the daylight saving time stretching from end of March to end of October.

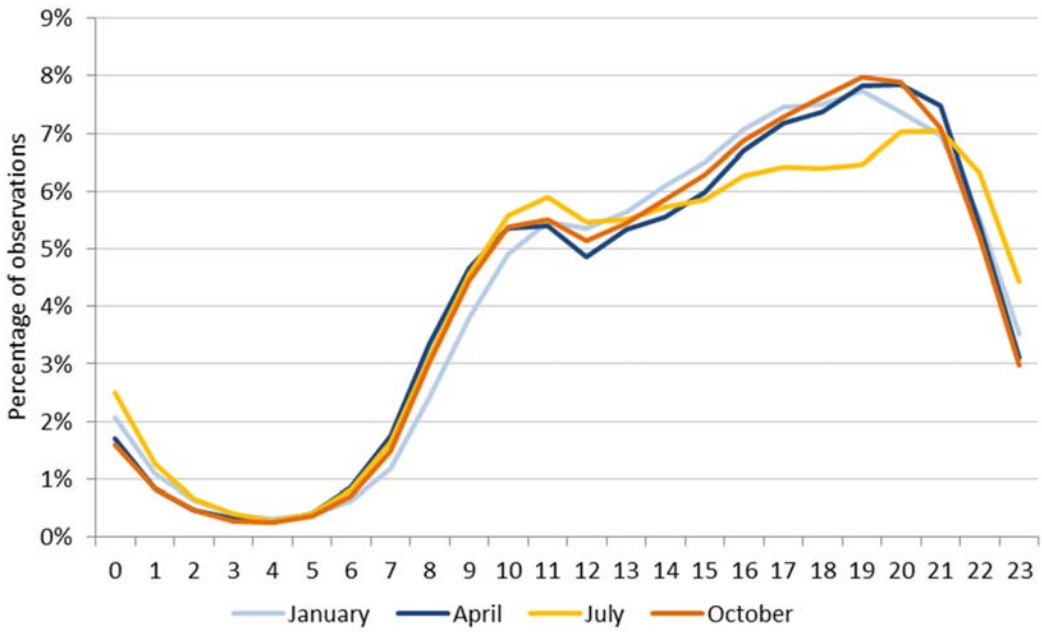
The result shows that most observations occur during the winter months with a seasonal drop during the summer months. This drop is expected considering the summer period is the time when most people in Sweden are on vacation. Figure 13 shows the share of Bredbandskollen observations by month for selected years.

Figure 13: Share of Bredbandskollen observations by month (selected years)



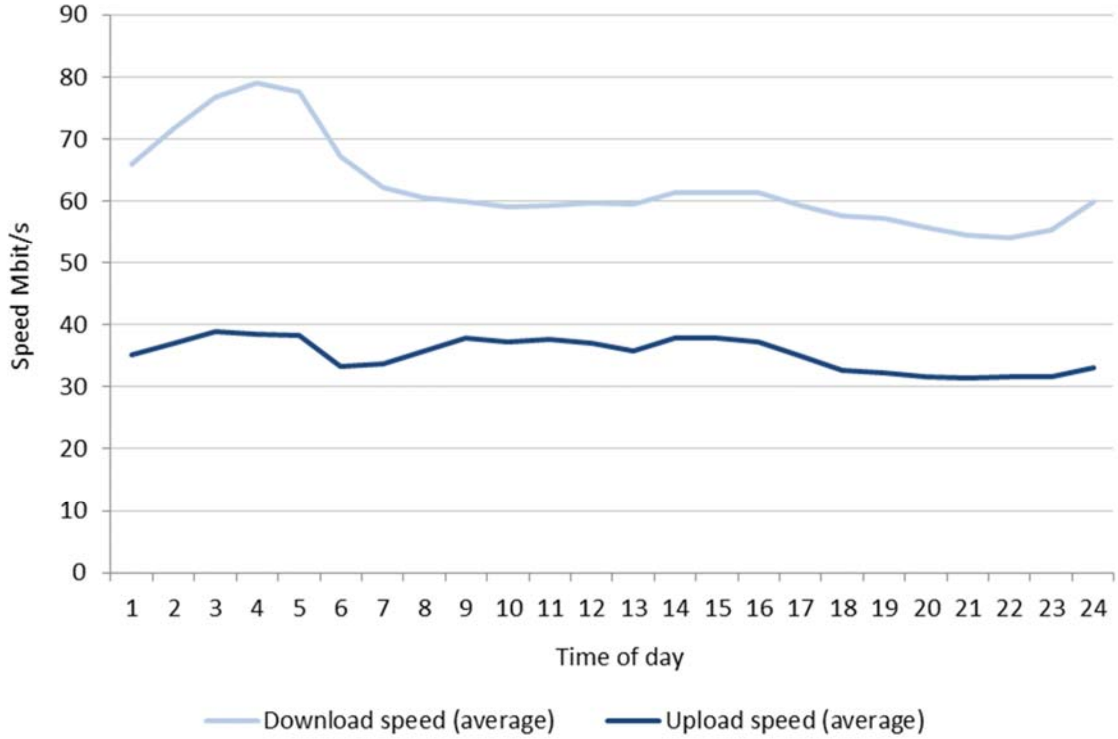
Patterns are also observed throughout the day. The share of all observations rises sharply throughout the morning up to around 11am. Following a short dip around 12am, which coincides with lunch time, the share of observations continues to rise throughout the afternoon with a peak around 6pm. This pattern is observed throughout the year except for the summer when the peak time is later in the evening and a greater share of observations is made during the night. Figure 14 shows the share of Bredbandskollen observations by time of day for selected months.

Figure 14: Share of Bredbandskollen observations by time of day (2016; selected months)



The average download speed is correlated with the share of observations, which in itself can be considered a sign of network traffic. As traffic decreases throughout the night, the average speed increases. However, despite an increasing share of observations throughout the day, the average download speed remains relatively constant throughout the day before again rising when people go to sleep. The average upload speed, however, is fairly constant throughout day and night. Figure 15 shows the average download and upload speed by hour.

Figure 15: Fixed-broadband average download and upload speeds by time of day, 2016)



4.4. ANALYSIS BY OPERATOR

Figure 16 shows the breakdown of observations per operator. When compared to the actual market shares (Swedish Post and Telecom Authority, 2016), there are some significant differences. The incumbent operator, Telia, is underrepresented, as well as Com Hem. On the other hand, Bahnhof and smaller operators included in the category others are largely overrepresented. Figure 17 shows the comparison between the share of Bredbandskollen observations per operator and the actual market shares.

Figure 16: Share of Bredbandskollen observations per operator.

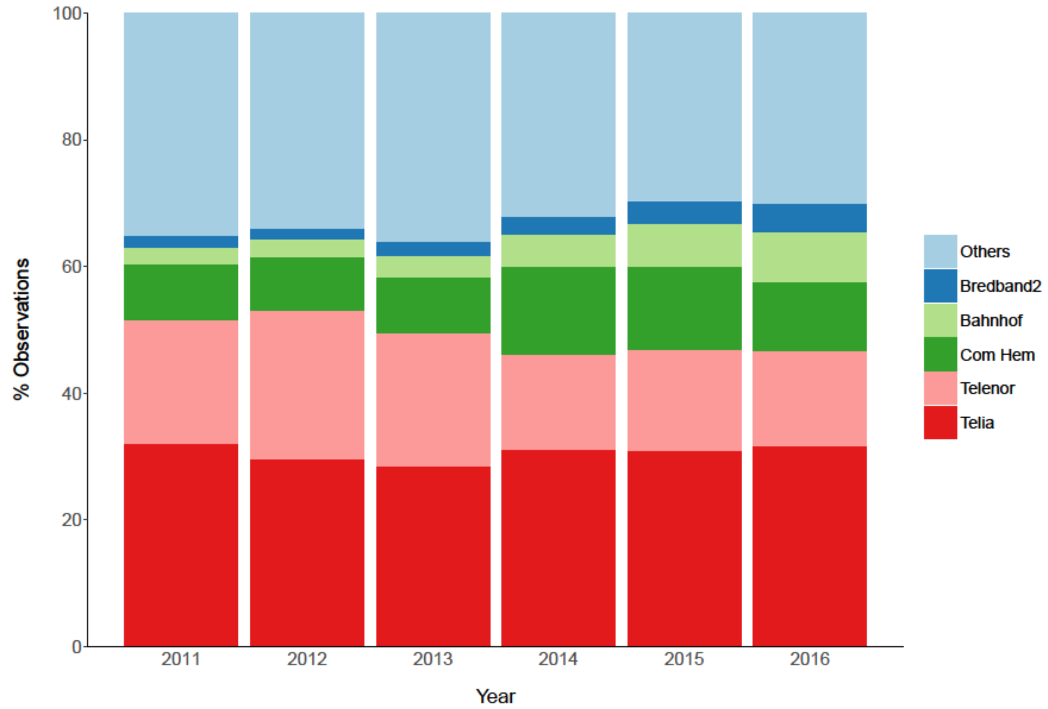
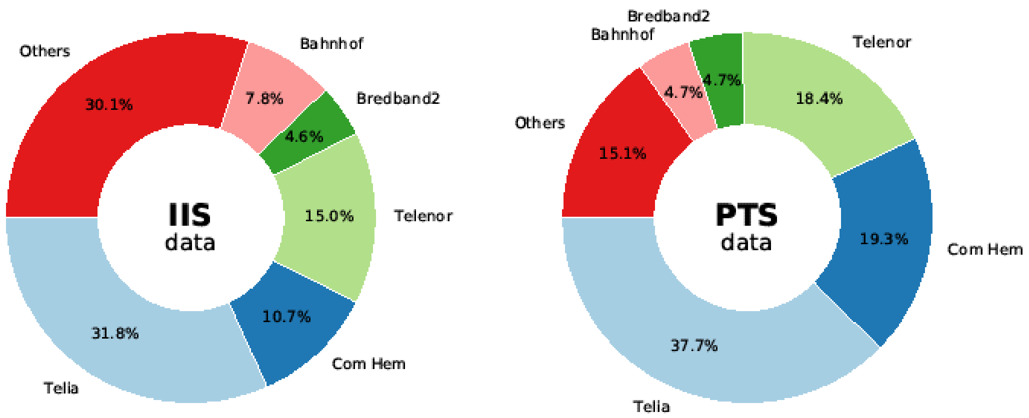


Figure 17: Share of Bredbandskollen observations per operator (left) and actual market shares (right), 2016.



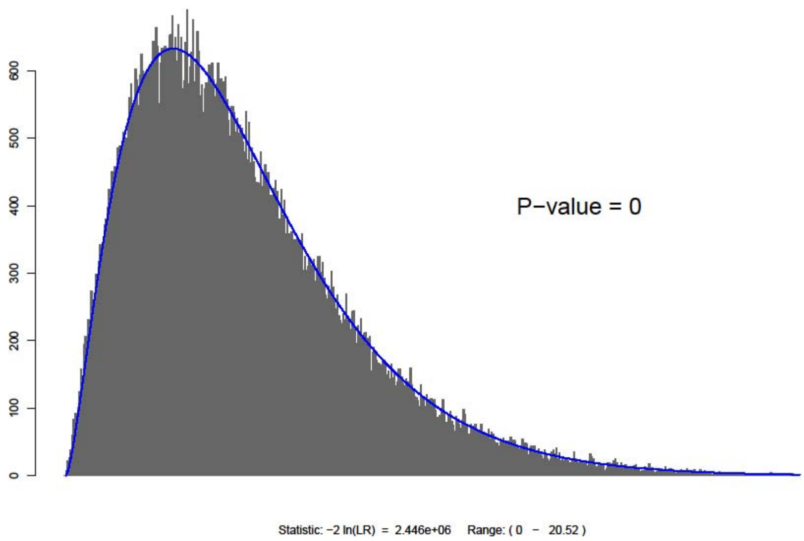
Source: Swedish Post and Telecom Authority (PTS) for market shares.

If the population represented in Bredbandskollen corresponded to the whole population of fixed-broadband subscriber in Sweden, the number of tests per operator should follow a multinomial distribution with the probability of each operator being equal to the market share. That is, if there were no selection bias, each test would be drawn (with replacement) from this multinomial distribution.

This formulation allows us to test the null hypothesis that Bredbandskollen's population is the same as that of all fixed-broadband subscribers in Sweden. If confirmed, this will be an indication of no (or mild) selection bias. If rejected, there will be evidence of an important selection bias.

Figure 18 shows the results of a chi-squared test, which strongly rejects the null. As could be expected given the large number of observations, the deviations between Bredbandskollen's share of observations per operator and the actual market shares indicate a significant selection bias.

Figure 18: Chi-squared test: share of Bredbandskollen observations per operator and actual market shares.

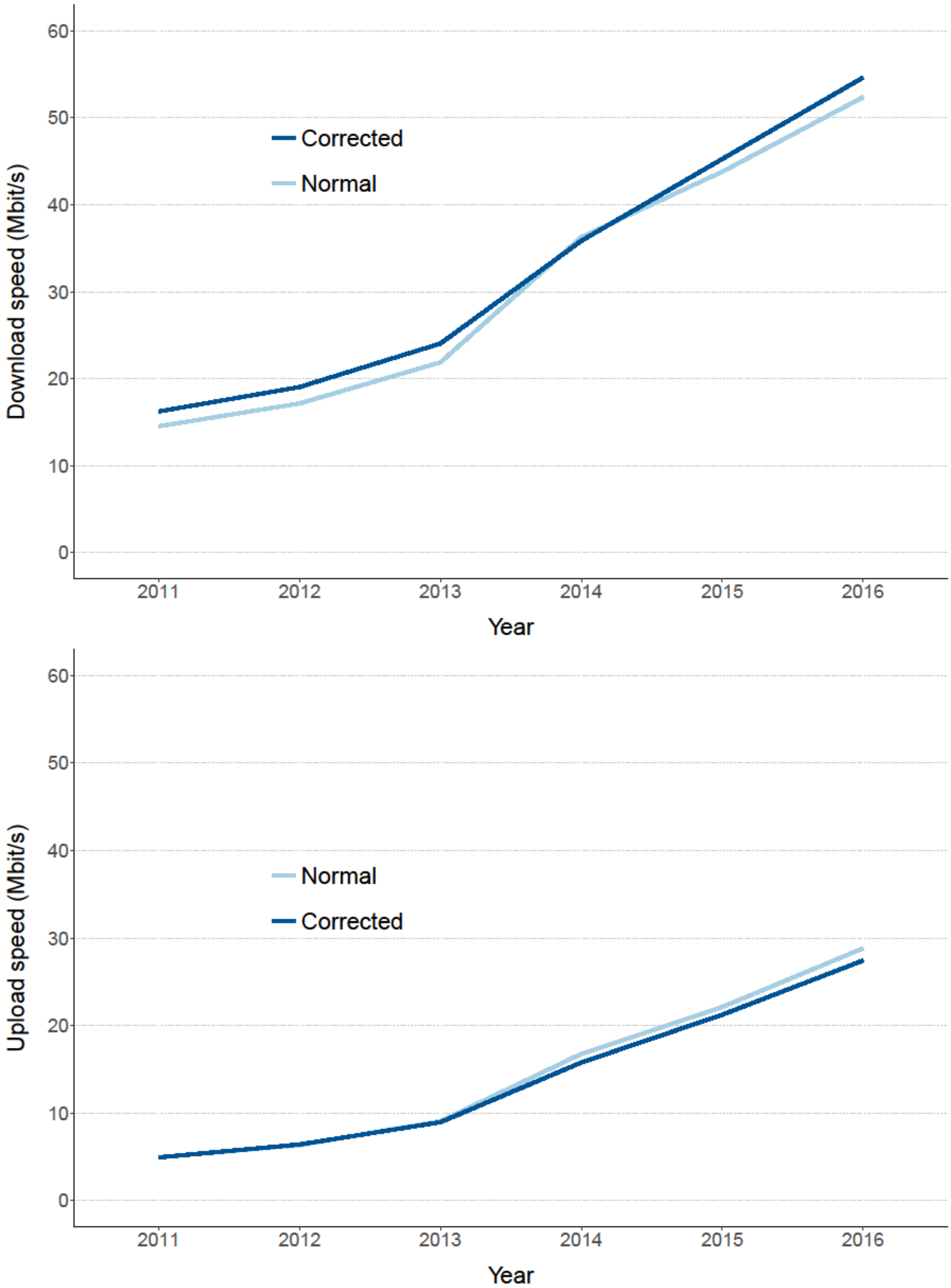


Source: Swedish Post and Telecom Authority (PTS) for market shares.

Post-stratification methods have been shown to eliminate selection bias in some non-representative samples (Zagheni and Weber, 2015). In particular, "post-stratification eliminates the bias due to selection or coverage problems if, within each adjustment cell, the probability that each case completes the survey is unrelated to the case value on the survey variable of interest" (Baker et al., 2013). In the case of broadband speed tests, this would mean that the end users of an under- or over-represented operator that decided not to perform the test did not do so because their performance would have been different from those end users from the same operator that performed the test. This is a strong assumption in our case.

Nevertheless, we test the effects of adjusting the weights given to each Bredbandskollen's observation so that the final distribution corresponds to actual market shares (Figure 19). The correction has, however, a minimal effect on the aggregate values.

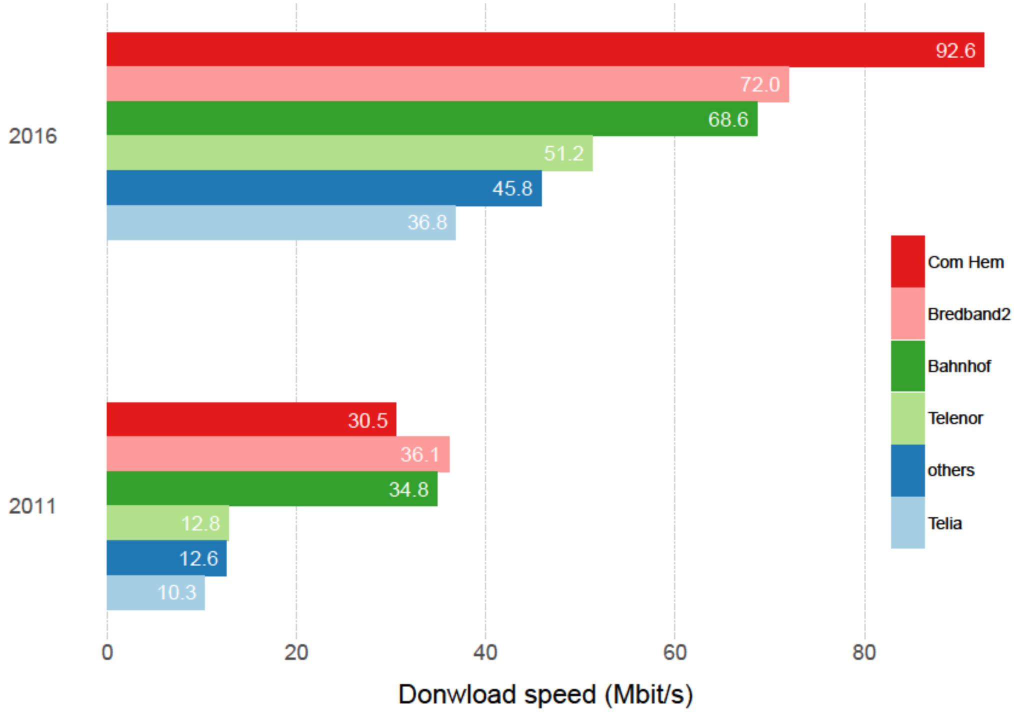
Figure 19: Fixed-broadband download speeds (top) and upload speeds (bottom), Sweden, 2011-2016 – Bredbandskollen adjusted test results.

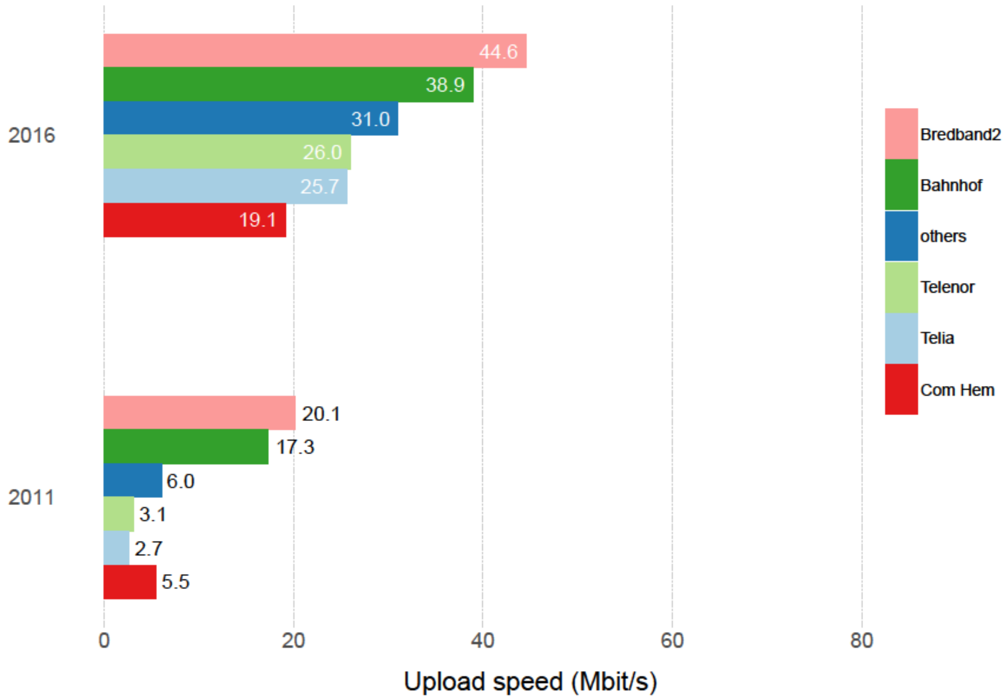


The analysis of Bredbandskollen's speed tests broken down by operator shows that although major differences exist between average speeds of different operators, they have opposite effects on the adjusted total average (Figure 20). That is, small operators ("others") and Bahnhof are over-represented and therefore their weight is adjusted downwards in the reweighted average. Conversely, Telia and Com

Hem are under-represented and their weight is adjusted upwards. The resulting adjusted average, however, barely changes.

Figure 20: Fixed-broadband download speeds (top) and upload speeds (bottom), by operator, Sweden 2011 and 2016 – Bredbandskollen test results.





5. CONCLUSION

The analysis of the data from Bredbandskollen carried out in this project has provided strong evidence of a selection bias. The lack of sufficient ground truth data on broadband speeds in Sweden does not allow to correct Bredbandskollen's sample in a way to ensure its representativity for the whole population of Sweden.

If Bredbandskollen could share part of the information it has on the number of unique users running the tests, some post-stratification adjustments could be envisaged to improve the representativeness of the sample. Even if the information on unique test users was shared at an aggregate level to avoid any privacy concerns (e.g. per region), that would enable a wider range of tests and adjustments to tackle the issue of the selection bias.

Another piece of information that would help in the correction of the bias is the breakdown of observations per technology (i.e. xDSL, fiber and coaxial cable). Bredbandskollen obtains information on the technology directly from network operators, in parallel to the test. This information is not contained in the log records shared with ITU, but would be very helpful. For instance, it would allow the linking of Bredbandskollen dataset with the statistics from the Swedish regulator on fixed-broadband technologies, thus adding another dimension from which to test and correct the selection bias.

Concerning the main findings of the analysis, it can be highlighted that there is a relative equality of speeds in all regions, supported by a very high baseline: 40 Mbit/s download and 20 Mbit/s upload. Even if the accuracy of the average speed measurement cannot be guaranteed given the evidence of a selection bias, the result that there are small differences between regions can still hold under the assumption that the selection bias is constant or evolves in parallel with time for all regions.

This would confirm that Sweden is an outstanding country in terms of inclusive high-speed broadband deployments.

6. REFERENCES

Baker, R., Brick, J., Bates, N., Battaglia, M., Couper, M., Denver, J., Gile, K., and Tourangeau, R. (2013). Non-probability sampling. http://www.aapor.org/Education-Resources/Reports/Non-Probability-Sampling.aspx. Report of the AAPOR Task Force, American Association for Public Opinion Research, Boston, MA.

Bauer, S., Clark, D. D., and Lehr, W. (2010). Understanding broadband speed measurements. https://groups.csail.mit.edu/ana/Publications/Understanding_broadband_speed_measurements_bauer_clark_lehr_TPRC_2010.pdf.

Bauer, S., Lehr, W., and Mou, M. (2016). Improving the measurement and analysis of gigabit broadband networks. In TPRC44, September 2016, Alexandria, VA.
Boletín Oficial del Estado, number 156, Friday 27 June (2014). https://www.boe.es/buscar/doc.php?id=BOE-A-2014-6729. Pages 49561-49583.

Boyd, D. and Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15(5):662–679.

Canadi, I., Barford, P., and Sommers, J. (2012). Revisiting broadband performance. In Proceedings of the 2012 ACM conference on Internet measurement, pages 273–286. ACM.

Chetty, M., Sundaresan, S., Muckaden, S., Feamster, N., and Calandro, E. (2013). Measuring broadband performance in south africa. In Proceedings of the 4th Annual Symposium on Computing for Development, page 1. ACM.

Couper, M. (2013). Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys. Survey Research Methods, 7(3):145–156.

Davidsson, P. (2017). Bredbandskollen Surfhastighet i Sverige 2008-2016. Technical report, Internetstiftelsen i Sverige (IIS).

Davies, R. (2016). Broadband as a universal service. Technical report, European Parliamentary Research Service. PE 581.977. Economides, N. (2008). Public Policy in Network Industries, chapter Handbook of Antitrust Economics. MIT Press, Cambridge, Mass.

European Commission (2010). A Digital Agenda for Europe. http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52010DC0245R%2801%29. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM(2010) 245 final/2.

European Telecommunications Standards Institute (2005). Speech Processing, Transmission and Quality Aspects (STQ); User related QoS parameter definitions and measurements; Part 4: Internet access. http://www.etsi.org/deliver/etsi_eg/202000_202099/20205704/01.01.01_60/eg_20205704v010101p.pdf. ETSI EG 202 057-4 V1.1.1 (2005-10).

Federal Communications Commission (2015a). 2015 Broadband Progress Report and Notice of Inquiry on Immediate Action to Accelerate Deployment. https://apps.fcc.gov/edocs_public/attachmatch/FCC-15-10A1_Rcd.pdf.

Federal Communications Commission (2015b). 2015 Measuring Broadband America Fixed Report. https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-broadband-america-2015.

International Telecommunication Union (2011). Handbook for the Collection of Administrative Data on Telecommunications/ICT 2011. http://www.itu.int/en/ITU-D/Statistics/Pages/publications/handbook.aspx.

International Telecommunication Union (2014). Manual for Measuring ICT Access and Use by Households and Individuals, 2014. http://www.itu.int/en/ITU-D/Statistics/Pages/publications/manual2014.aspx.

International Telecommunication Union (2016). Measuring the Information Society Report 2016. http://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2016.aspx.

Lehr, W., Bauer, S., and Clark, D. D. (2013). Measuring Performance when Broadband is the New PSTN. Journal of Information Policy, 3:411–441. M-Lab Research Team and others (2014). ISP interconnection and its impact on consumer Internet performance – a measurement lab consortium technical report. https://www.measurementlab.net/publications/M-Lab_Interconnection_Study_US.pdf.

Ministerio de Energía, Turismo y Agenda Digital (2017). Informe de seguimiento de los niveles de calidad de servicio. Parámetros específicos para el servicio de acceso a Internet: Velocidad de transmisión de datos conseguida. Primer trimestre de 2017. http://www.minetad.gob.es/telecomunicaciones/es-ES/Servicios/CalidadServicio/informes/Documents/Seguimiento_SAI_T1_17.pdf.

OECD (2011). OECD Regional Typology, 2011. https://www.oecd.org/cfe/regional-policy/OECD_regional_typology_Nov2012.pdf

Ofcom (2017). UK fixed-line broadband performance, November 2016: Research Report. https://www.ofcom.org.uk/research-and-data/telecoms-research/broadband-research/uk-home-broadband-performance-2016.

Prasad, T., Shekhawat, D., Guha, S., Goveas, N., and Deshpande, B. (2016). Analysis of impartial quality measurements on indian broadband connections. In Communication (NCC), 2016 Twenty Second National Conference on, pages 1–6. IEEE.

Riddlesden, D. and Singleton, A. D. (2014). Broadband speed equity: A new digital divide? Applied Geography, 52:25–33.

Rohman, I. K. and Bohlin, E. (2012). Does broadband speed really matter as a driver of economic growth? investigating oecd countries. International Journal of Management and Network Economics, 2(4):336–356. PMID: 51888.

Rood, H., Yoshikawa, D., Wevers, R., Post, R., Tetteroo, A.-J., Kuipers, A., and Schurmann, R. (2012). Usability of broadband performance measurements for statistical surveys. In 2012 TRPC.

SamKnows (2013). Quality of Broadband Services in the EU (March 2012). Technical report, European Commission.

SamKnows (2014). Quality of Broadband Services in the EU (October 2013). Technical report, European Commission.

Sundaresan, S., Burnett, S., Feamster, N., and De Donato, W. (2014). BISmark: A Testbed for Deploying Measurements and Applications in Broadband Access Networks. In USENIX Annual Technical Conference, pages 383–394.

Sundaresan, S., De Donato, W., Feamster, N., Teixeira, R., Crawford, S., and Pescape, A. (2012). Measuring home broadband performance. Communications of the ACM, 55(11):100–109.

Sundaresan, S., Feamster, N., and Teixeira, R. (2016). Home Network or Access Link? Locating Last-Mile Downstream Throughput Bottlenecks. In PAM 2016 - Passive and Active Measurement Conference, pages 111–123, Heraklion, Greece.

Swedish Post and Telecom Authority (2016). The Swedish Telecommunications Market – First Half-year 2016. https://www.pts.se/upload/Rapporter/Tele/2016/Swedish-Telekommunications-Market-first-half-2016.pdf.

Telecom Regulatory Authority of India (2006). Regulation quality of service standards for broadband services.

Telecom Regulatory Authority of India (2014). Quality of service of broadband service (second amendment) regulations. http://www.trai.gov.in/sites/default/files/201406250411117962118scan0004_0.pdf.

Tiru, M. (2016). Big Data for Measuring the Information Society. In 14th World Telecommunication/ICT Indicators Symposium, Botswana.

TNS Opinion & Social (2014). Special Eurobarometer 414 "E-Communications and Telecom Single Market Household Survey". Technical report, European Commission.

United Nations Economic and Social Council (2016). Statistical Commission, Fortyseventh session, Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators. https://unstats.un.org/unsd/statcom/47th-session/documents/2016-2-IAEG-SDGs-Rev1-E.pdf. E /CN.3/2016/2/Rev.1 (8-11 March 2016).

Wattegama, C. and Kapugama, N. (2011). Volunteer computing model prospects in performance data gathering for broadband policy formulation. Communications and Strategies, (81):153–174.

Whitacre, B., Gallardo, R., and Strover, S. (2014). Broadband's contribution to economic growth in rural areas: Moving towards a causal relationship. Telecommunications Policy, 38(11):1011–1023.

Zagheni, E. and Weber, I. (2015). Demographic research with non-representative internet data. International Journal of Manpower, 36(1):13–25.

Zuhyle, S. and Mirandilla-Santos, G. (2015). Measuring broadband performance: Lessons learnt, challenges faced. In Proceedings of CPRSouth 2015, Taiwan.