



Ipsos MORI
Social Research Institute

October 2020

EIB Group Survey of Investment and Investment Finance

Technical report

Contents

Glossary	2
Member State codes	2
1 Introduction	3
1.1 Purpose and structure of the report	3
1.2 Survey objectives	3
1.3 Overview of methodology	3
2 Sampling and weighting	11
2.1 Survey population and requirements	11
2.2 Creation of the sampling frame for the survey	12
2.3 Sample selection and stratification	14
2.4 Weighting	15
3 Questionnaire translation	21
4 Fieldwork	23
4.1 Interviewer briefings and training	23
4.2 Fieldwork supporting materials	23
4.3 Fieldwork platform	24
4.4 Mainstage fieldwork period	24
4.5 Weekly fieldwork reporting	26
4.6 Response rates	27
4.7 Consent to take part in future General Module survey	31
5 Quality control and data processing	32
5.1 Quality control pre-fieldwork	32
5.2 Quality control during fieldwork	32
5.3 Quality control post fieldwork	35
(a) Data validation – Routing	36
(b) Data validation – Permitted values	36
(c) Descriptives – Impossible and implausible responses	37
(d) Data validation – Duplicates and near duplicates observations	38
(e) Item non-response	40
5.4 Back-coding	42
5.5 Data linking	42
6 Appendix	43

Glossary

Member State codes

AT	Austria	IE	Ireland
BE	Belgium	IT	Italy
BG	Bulgaria	LT	Lithuania
CY	Cyprus	LU	Luxembourg
CZ	Czech Republic	LV	Latvia
DE	Germany	MT	Malta
DK	Denmark	NL	Netherlands
EE	Estonia	PL	Poland
EL	Greece	PT	Portugal
ES	Spain	RO	Romania
FI	Finland	SE	Sweden
FR	France	SI	Slovenia
HR	Croatia	SK	Slovakia
HU	Hungary	UK	United Kingdom
		US	United States

1 Introduction

1.1 Purpose and structure of the report

This “Technical report” for the fifth wave of the EIB Group Survey on Investment and Investment Finance (EIBIS) covers the detail of design and implementation of all constituent surveys of this study: the General Module, Add-on Module (surveying municipalities) and the Online Module. The chapters of this report consider the overall objectives and methodology of the survey(s), sampling and weighting, questionnaire development and translation, fieldwork execution, and quality control and data processing.

1.2 Survey objectives

The EIBIS is carried out to provide the European Investment Bank (EIB) Group with robust and comparable evidence on the corporate investment environment across all 27 European Union (EU) countries, the UK¹ and the United States of America². Data has been collected via five annual surveys (or waves) in 2016, 2017, 2018, 2019 and 2020. In the first 4 waves (2016, 2017, 2018 and 2019³) the survey consisted of: 1) a General Module survey; 2) an Add-on Module survey; and 3) an Online Follow-up Module survey. In the 5th Wave (2020) the survey consisted of 1) a General Module survey; 2) an Add-on Module survey; and 3) a stand alone Online Module survey. The General Module was similar for all five waves and collected time series and longitudinal data on investment and investment finance decisions. The Add-on Module changes year-on-year to provide EIB with the flexibility to respond to new priorities. In year 2020 the Add-on Module was similar to the one in 2017. It functioned as a ‘Municipality survey’ and examined the infrastructure investment needs in municipalities across the EU⁴, with a specific focus on the impact of climate change on those investment needs. The Online Module was a stand alone survey this year which aimed to provide more detailed understanding of the investment decisions of businesses in supporting climate action and reducing carbon emissions.

This report covers the fifth wave of the survey which was conducted in the spring/summer of 2020.

1.3 Overview of methodology

1.3.1 General Module methodology

The aim of the General Module survey was to provide the EIB Group with insights on market trends in investment finance and the drivers of investment decisions among the non-financial corporate (NFC) sector in all 27 EU Member States, the UK and the US. The US market provides direct comparison to the EU. The 2020 General Module Survey was administered to senior managers or financial directors in 13,372 establishments.

Fieldwork for the 2020 wave of the General Module started in April 2020 (see more details in the Fieldwork chapter) and for the majority of countries was completed by the end of July, a couple of countries continued into early August. The

¹ The United Kingdom (UK), is excluded from the EU sample as it was no longer an EU Member State for the EIBIS 2020 fieldwork and reporting period.

² The US was included in 2019 and 2020 to act as a comparison group for the EU sample.

³ There were two Add-on modules in 2019 – (1) Measuring the satisfaction with EIB among clients and potential clients, and (2) Gaining a better understanding of start-up and scale-up businesses.

⁴ The Add-on module is conducted only in the EU countries (i.e. excluding the UK and the US).

survey was administered by telephone with a median interview length of 28 minutes, which includes three minutes of screening questions to check that sampled businesses were eligible for the survey.

1.3.2 Sampling methodology General Module

The sample for the General Module represents non-financial enterprises⁵ in the 27 EU Member States, the UK and the US, in NACE categories C to J with a minimum of five employees⁶. Eligible respondents were senior persons with responsibility for investment decisions and how these are financed. This person could be the owner, a Finance Manager, the Finance Director or Head of Accounts, the Chief Financial Officer (CFO) or the Chief Executive Officer (CEO). For the General Module Survey, the Bureau van Dijk ORBIS dataset was used as the sampling frame in all countries. Given that ORBIS does not include telephone numbers for approximately 25% of the survey population, a telephone number search was undertaken using Dun & Bradstreet.

The sample for the General Module included a panel component as well as the top-up sample. Panel firms were those that participated in the previous wave of the survey, and consented to be re-contacted in the following wave. All firms that had consented to be re-contacted in wave 4 were included in the panel sample; and they were interviewed even if they mentioned an ineligible size or sector category during the screening questions (due to having previously confirmed eligibility). The top-up sample, on the other hand, included firms that did not participate in wave 4 (though in some instances they may have taken part in the first three waves – either with the same or a different interviewee).

The method adopted for selecting a top-up sample from ORBIS was random stratified sampling. This was the same approach used in previous waves (conducted in 2016, 2017, 2018 and 2019). The sample was stratified disproportionately by country, industry group and size class, and stratified proportionally by region. Where possible an initial sample ratio of 8:1 (post telephone matching) was used – that is, eight cases sampled for each achieved interview. However, this often proved insufficient in waves 1 to 4, so the sample ratio was re-adjusted for some countries taking into consideration previous call outcomes. A reserve sample was also prepared and used in many countries to compensate for issues with sample quality and/or availability of respondents during fieldwork. If necessary, an additional reserve sample was prepared and uploaded for countries with low fieldwork efficiency. For more details on the sampling, see section 2.

1.3.3 Sample size targets General Module

The achieved country level sample sizes for the 2020 General Module Survey were as per Table 1 below, the targets are shown in brackets. All country level sample targets for wave 5 were achieved, except in Latvia which had difficulty recruiting medium-size and large businesses and exhausted the sample for these enterprises, finishing on 389 interviews (instead of 400).

⁵ The sampling targeted head offices; branches were filtered out of the sampling frame, but any branches or sites were not screened out in the interview. Given the length of the screening questionnaire and low incidence of such firms, it was decided not to screen out branches in the interview during fieldwork.

⁶ The survey covered NACE C (manufacturing), NACE G & I (services), NACE F (construction) and NACE D & E & H & J (infrastructure). It included freelancers working regularly for a company. Full-time and part-time employees were counted as one employee. Employees working less than 12 hours per week were excluded.

Table 1: General Module completes and targets, by country⁷

Country	N	Target	Country	N	Target	Country	N	Target
AT	480	(480)	FI	480	(480)	NL	480	(480)
BE	480	(480)	FR	601	(600)	PL	483	(480)
BG	480	(480)	HR	488	(480)	PT	481	(480)
CY	180	(180)	HU	481	(480)	RO	480	(480)
CZ	481	(480)	IE	401	(400)	SE	480	(480)
DE	601	(600)	IT	601	(600)	SI	400	(400)
DK	480	(480)	LT	400	(400)	SK	400	(400)
EE	400	(400)	LU	180	(180)	UK	601	(600)
EL	403	(400)	LV	370	(400)	US	800	(800)
ES	600	(600)	MT	180	(180)	Total	13372	13,380

Quotas (sample size targets) were again set within each EU Member State and the United States on four industry groupings: (i) manufacturing – NACE C, (ii) services – NACE G & I, (iii) construction - NACE F, and (iv) infrastructure – NACE D & E & H & J; and four size-classes: (i) micro (5-9); (ii) small (10-49); (iii) medium (50-249); and (iv) large (250+).

As in previous waves, wherever country sample sizes allowed for it, indicative quotas were set to achieve a minimum sample size of c.100 interviews in each group. In practice this meant that:

- In countries with 600 interviews, quotas of approximately 150 were set in each group
- In countries with 480 interviews, quotas of approximately 120 were set in each group
- For the smaller country samples – of 400 or 180 – the sample was split evenly across the categories (100 in each or c.45 in each)
- In the United States – 800 interviews, quotas of approximately 200 were set in each group

Table 2 below shows the total number of achieved interviews ('achieved' columns) per size and sector in each country, next to the set quotas. The original sector and size quota targets were relaxed for some countries considering the achieved sample sizes in waves 1 to 4 (plus available sample in some smaller countries).

Table 2: Quotas – targets and achieved, per sector, size and country

		Target	Achieved	Target	Achieved	Target	Achieved	Target	Achieved	Achieved	Target	Achieved
	Size:	micro (5-9)	micro (5-9)	small (10-49)	small (10-49)	medium (50-249)	medium (50-249)	large (250+)	large (250+)		Total	Total
	Sector:	manufacturing – NACE C	manufacturing – NACE C	services – NACE G & I	services – NACE G & I	construction - NACE F	construction - NACE F	infrastructure – NACE D & E & H & J	infrastructure – NACE D & E & H & J	Ineligible*	Total	Total
AT	Size	120	90	120	128	120	128	120	134	0	480	480
	Sector	120	147	120	121	120	86	120	120	6	480	480
BE	Size	120	93	120	155	140	168	100	64	0	480	480
	Sector	120	150	120	118	120	92	120	107	13	480	480
BG	Size	120	119	120	130	140	148	100	83	0	480	480
	Sector	120	140	120	115	120	115	120	108	2	480	480
CY	Size	40	42	75	87	50	41	15	10	0	180	180
	Sector	50	58	50	83	40	25	40	13	1	180	180
CZ	Size	120	108	130	144	130	139	100	90	0	480	481
	Sector	120	161	120	122	120	91	120	107	0	480	481
DE	Size	150	143	150	204	150	157	150	97	0	600	601
	Sector	150	167	150	150	150	141	150	134	9	600	601
DK	Size	80	86	160	165	160	161	80	68	0	480	480
	Sector	130	151	120	108	100	106	130	115	0	480	480
EE	Size	100	148	160	144	120	97	20	11	0	400	400
	Sector	100	110	100	99	100	92	100	99	0	400	400
EL	Size	100	88	100	117	100	123	100	75	0	400	403
	Sector	100	118	100	116	100	86	100	77	6	400	403
ES	Size	150	130	150	173	150	175	150	122	0	600	600
	Sector	150	170	150	147	150	138	150	142	3	600	600
FI	Size	120	129	120	150	160	135	80	66	0	480	480
	Sector	120	125	120	116	120	119	120	112	8	480	480
FR	Size	150	162	150	187	150	160	150	92	0	600	601
	Sector	150	177	150	145	150	139	150	135	5	600	601
HR	Size	120	82	120	169	150	164	90	73	0	480	488
	Sector	120	149	120	108	120	97	120	127	7	480	488

Table 2: Quotas – targets and achieved, per sector, size and country

		Target	Achieved	Target	Achieved	Target	Achieved	Target	Achieved	Achieved	Target	Achieved
	Size:	micro (5-9)	micro (5-9)	small (10-49)	small (10-49)	medium (50-249)	medium (50-249)	large (250+)	large (250+)		Total	Total
	Sector:	manufacturing – NACE C	manufacturing – NACE C	services – NACE G & I	services – NACE G & I	construction - NACE F	construction - NACE F	infrastructure – NACE D & E & H & J	infrastructure – NACE D & E & H & J	Ineligible*	Total	Total
HU	Size	120	118	120	157	150	158	90	48	0	480	481
	Sector	120	145	120	110	120	122	120	99	5	480	481
IE	Size	100	131	160	222	120	44	20	4	0	400	401
	Sector	100	110	100	148	100	66	100	65	12	400	401
IT	Size	150	118	150	204	150	175	150	104	0	600	601
	Sector	150	207	150	157	150	130	150	97	10	600	601
LT	Size	100	103	120	124	150	124	30	49	0	400	400
	Sector	100	112	100	93	100	85	100	102	8	400	400
LU	Size	40	24	75	72	40	61	25	23	0	180	180
	Sector	45	46	45	51	45	43	45	38	2	180	180
LV	Size	100	84	130	174	140	90	30	22	0	400	370
	Sector	100	98	100	91	100	77	100	99	5	400	370
MT	Size	20	25	90	87	55	52	15	16	0	180	180
	Sector	40	44	80	73	25	16	35	45	2	180	180
NL	Size	80	86	150	145	150	161	100	88	0	480	480
	Sector	120	140	120	119	120	99	120	120	2	480	480
PL	Size	80	37	150	158	150	186	100	102	0	480	483
	Sector	120	141	120	108	120	89	120	141	4	480	483
PT	Size	100	98	120	132	140	162	120	89	0	480	481
	Sector	120	150	120	116	120	96	120	119	0	480	481
RO	Size	120	110	120	159	140	142	100	69	0	480	480
	Sector	120	139	120	103	120	112	120	124	2	480	480
SE	Size	120	109	120	143	140	146	100	82	0	480	480
	Sector	120	138	120	113	120	98	120	123	8	480	480
SI	Size	100	95	120	119	120	121	60	65	0	400	400
	Sector	100	125	100	96	100	96	100	81	2	400	400

Table 2: Quotas – targets and achieved, per sector, size and country

		Target	Achieved	Target	Achieved	Target	Achieved	Target	Achieved	Achieved	Target	Achieved
	Size:	micro (5-9)	micro (5-9)	small (10-49)	small (10-49)	medium (50-249)	medium (50-249)	large (250+)	large (250+)		Total	Total
	Sector:	manufacturing – NACE C	manufacturing – NACE C	services – NACE G & I	services – NACE G & I	construction - NACE F	construction - NACE F	infrastructure – NACE D & E & H & J	infrastructure – NACE D & E & H & J	Ineligible*	Total	Total
SK	Size	100	87	100	135	120	128	80	50	0	400	400
	Sector	100	130	100	103	100	77	100	90	0	400	400
UK	Size	150	135	150	214	150	195	150	57	0	600	601
	Sector	150	171	150	186	150	130	150	108	6	600	601
US	Size	200	181	200	261	200	234	200	124	0	800	800
	Sector	200	215	200	222	200	180	200	171	12	800	800
Total	Size	3,170	2961	3,750	4459	3,835	3975	2,625	1977	0	13,380	13372
	Sector	3,355	3934	3,385	3437	3,300	2843	3,340	3018	140	13,380	13372

*"Ineligibles" are some panel firms from wave 4. Panel firms are interviewed even if they say they do not now meet the sector and size requirements. Those that say they now operate in a NACE sector that is not C-J are excluded from the sub-group analysis/reporting, but those firms that say they now have less than 5 employees are considered as micro companies in sub-group analysis/reporting.

1.3.4 Add-on Module methodology

In 2020, the study included an additional module which focused on infrastructure investment needs of municipalities. The survey targeted local government municipalities to better understand their infrastructure investment needs, to help the EIB Group improve and extend the reach of its products and services, better tailoring them to the needs of municipalities. This survey aimed to interview the most senior person in municipalities with responsibility for infrastructure investment decisions and how these are financed (e.g., mayor, treasurer or chief civil engineer).

The Add-on Module was administered by telephone and took an average (median) of 40 minutes to complete. Mainstage fieldwork took place between May and July 2020. As the Add-on Module did not change substantially after the pilot, 83 pilot interviews were retained in the final data set.

The Add-on Module aimed to interview respondents in at least 650 municipalities. EIB devised a sampling strategy based on achieving a given number of interviews in each NUTS2 region with a focus on achieving interviews with the larger municipalities.

1.3.5 Sampling methodology Add-on Module

The municipality sample was prepared by EIB using different sources of information including ORBIS, European Commission, and Internet searches. Ipsos reviewed and improved the sample where necessary by cleaning/formatting phone numbers, removing duplicates, and ordering phone numbers. Ipsos country offices also reviewed and updated the sample files to reflect most up to date contact names and correct phone numbers. An overall response rate of 31% was achieved (see also section on response rates in Fieldwork chapter below).

The Add-on Module targets were met across nearly all countries, and exceeded in many countries. Table 3 below provides an overview of the achieved Add-on Module interviews per country against initial targets.

<i>Country</i>	<i>N</i>	<i>Target</i>	<i>Country</i>	<i>N</i>	<i>Target</i>	<i>Country</i>	<i>N</i>	<i>Target</i>
<i>AT</i>	17	(17)	<i>FI</i>	18	(18)	<i>NL</i>	41	(41)
<i>BE</i>	39	(41)	<i>FR</i>	47	(50)	<i>PL</i>	50	(50)
<i>BG</i>	27	(27)	<i>HR</i>	20	(20)	<i>PT</i>	17	(17)
<i>CY</i>	5	(5)	<i>HU</i>	27	(27)	<i>RO</i>	45	(45)
<i>CZ</i>	28	(27)	<i>IE</i>	7	(5)	<i>SE</i>	27	(27)
<i>DE</i>	50	(50)	<i>IT</i>	57	(50)	<i>SI</i>	7	(7)
<i>DK</i>	18	(18)	<i>LT</i>	14	(14)	<i>SK</i>	14	(14)
<i>EE</i>	9	(9)	<i>LU</i>	4	(4)			
<i>EL</i>	28	(28)	<i>LV</i>	14	(14)			
<i>ES</i>	50	(50)	<i>MT</i>	5	(5)	<i>Total</i>	685	(670)

1.3.6 Online Module methodology

An online survey was devised to understand the investment decisions of businesses in supporting climate action and reducing carbon emissions. The sample for the survey was sourced from Crunchbase and ORBIS. All firms on Crunchbase were considered eligible for the survey, whilst the ORBIS sample reflected the sample for the General Module, only including non-financial enterprises⁸ in the 27 EU Member States, the UK and the US, in NACE categories C to J with a minimum of five employees⁹ (although no firm was screened out of the survey if they had fewer than 5 employees). Only firms with an email address were selected for inclusion in the sample and all firms who had taken part in the General Module survey and had agreed to participate in the next wave of the General Module survey were excluded.

An email invitation was sent to all firms with an email address, the invitation provided details on the background to the survey, its aims and a description of whom the email should be forwarded to i.e. the type of job role who would best be able to answer the questions. The majority of email addresses on the databases were general email addresses rather than named individuals and therefore it was not possible to target relevant individuals.

To maximise participation, each email invitation was sent out with a personalised link to the online survey. In addition, two reminder emails were sent during the fieldwork period. The emails reflected whether the respondent had already clicked on the link to maximise the response rate.

In total, 1,034,785 respondents were invited to participate in the online module. Of these, 1,609 respondents completed the online module.

⁸ The sampling targeted head offices; branches were filtered out of the sampling frame, but any branches or sites were not screened out in the interview. Given the length of the screening questionnaire and low incidence of such firms, it was decided not to screen out branches in the interview during fieldwork.

⁹ The survey covered NACE C (manufacturing), NACE G & I (services), NACE F (construction) and NACE D & E & H & J (infrastructure)..

2 Sampling and weighting

This chapter describes the procedures for sampling and weighting for the General Module of the survey. This broadly followed the same approach as in previous years.

2.1 Survey population and requirements

As in previous years, the General Module sample was required to represent non-financial enterprises in the 27 EU Member States, the UK and US, with at least five employees, and belonging to one of NACE categories C to J. The US was included for the first time in 2019. The sample consisted of two components: the panel (longitudinal) sample (EU only) and a cross-sectional or “top-up” sample. Total sample sizes were targeted across the sum of these samples with variation depending on the size of the country, with a total sample of 13,380 enterprises, with 180, 400, 480 or 600 interviews per EU country and 800 interviews for the US (see table 1 in section 1.3.3).

The longitudinal sample issued for the survey included all firms that were interviewed in the previous year and agreed re-contact permission for future survey years. All firms in this sample were eligible for the survey even if they would otherwise be ineligible due to a size or sector change between years.

The cross-sectional sample was made up of cases from the latest ORBIS database, less the panel cases (i.e. those interviewed in the previous year who agreed re-contact permission) and those that had a previous year survey outcome of ‘discontinued business’, ‘not a business’ or ‘ineligible activity’. Firms interviewed in previous years but not the most recent year, and which therefore did not qualify to be part of the latest panel sample, were also in scope for selection in this part of the sample, and were treated in the same way as other cross-sectional cases. Firms in this sample had to meet the full survey eligibility requirements.

For the purposes of this survey an enterprise was defined as a company trading as its own legal entity. As branches were excluded from the population, the survey population is closer to an enterprise (head office) than an establishment level survey. However, the definition is broader than a typical enterprise survey given that some company subsidiaries are their own legal entities.

Separate estimates with similar levels of precision in each country were required for the following industry groupings and size classes:

Sector:

- manufacturing – NACE C,
- services – NACE G & I,
- construction - NACE F, and
- infrastructure – NACE D & E & H & J,

Size class:

- micro (5-9),
- small (10-49),
- medium (50-249), and
- large (250+).

As noted earlier, the sample was based on a quota sampling methodology, with equal-sized sample quotas set in each country on the industry grouping and size classes listed above. However, the sector and size quotas were again revised slightly in some countries considering response rates and the availability of sample in each quota cell using achieved quotas in previous waves as a benchmark (see Table 2 in section 1.3.3).

The sample selection procedures described in this section are based on what would be done if applying an equivalent random probability survey methodology; the difference came at the fieldwork stage when the fieldwork was conducted based on a quota methodology, with fieldwork teams instructed to deliver the interviews to meet the quota targets.

2.2 Creation of the sampling frame for the survey

As in previous years, Bureau van Dijk's ORBIS database was used to draw a sample of new cases to be included in the survey. The process of creating a sampling frame from which to select the cases is described in this section.

2.2.1 ORBIS selection criteria

The sample extraction required creating country-level sampling frames from the latest SQL ORBIS databases. It followed the same process as was used in previous years, with the end result a set of country-level files which contained the full population of eligible firms including all potentially eligible firms which did not contain information on the number of employees (so that this could be imputed).

The following process and selection criteria, matching the process developed with EIB in previous years, were applied to produce the files:

First, the contact information database was imported and split into 29 countries based on ISO country code. The required variables were kept from this initial import file. Following this, "the legal info" database was imported and matched onto the contact information on BVD ID number. The resulting firms were then filtered to include: those that were known to be active or where activity status was unknown; corporate firms; and to exclude branches and non-profit organisations. These steps ensured that the firms retained in the sampling frame were aligned with the survey population.

Subsequently, the "Industry – Global financials and ratios" database was imported and matched onto the set of cases from earlier steps. This included financial variables required for the sampling frame and imputations. The import was filtered on closing date (most recent), consolidation code (not C2) and, separately, number

of employees (nonblank). In cases of duplicates on the most recent closing date the selection was based on a hierarchy on consolidation code (U1, U2, C1 then LF). ORBIS generally contains multiple listings per company, one per reporting episode. The consolidation code filtering removes listings that pertain to holding companies, as these will tend to over-report the financials and numbers of employees of the individual enterprises. The step to retain number of employees separately from the most recent financial listing was done so that if the number of employees was not reported at the most recent listing the information could instead be used from an earlier listing, maximising the number of cases that had this information.

Finally, the “industry classifications” database was imported. This was used to attach the industry sector of each case, and subsequently drop cases which did not fit the survey population (i.e. those not in NACE section C-J). NACE top-level code was used for EU firms, however the US does not use this classification system, and so the NAICS 2012 classification was first derived into NACE classifications.

At the end of this process counts were output of the number of firms extracted in each country, split by size category (missing, out of scope (0-4), micro (5-9), small (10-49), medium (50-249) and large (250+)). Counts were also provided at the country level for each stage in the process to help verify the numbers. These were compared with the same counts from previous years, to verify that the extraction process (and ORBIS data) were reasonably consistent year on year. Most countries had stable numbers on survey population eligible establishments¹⁰.

2.2.2 Imputation approach

The purpose of the imputation was to provide accurate estimations of company sizes where missing in the ORBIS data set. This activity included two stages, the statistical work to impute missing size values, and then agreeing which cases to keep in the sampling frames. The imputations used the same statistical techniques used in previous waves.

The size categories of all of the cases in the sampling frame were imputed if there was any information available that could be used for the predictions. Cases with no information at all were not imputed. Imputation was undertaken even if size information was available from ORBIS. The numbers of firms were then compared, across size and sector categories, with external population data, for several potential solutions. Identifying these options generally involved setting an inclusion cut-off at different years of when the financials were reported (i.e. including cases where the financials were reported 3 years ago or more recently, 4 years ago, 5 years ago, etc.). The solutions which performed best and so were chosen were the same as in previous years¹¹. Specifically, in 26 countries the cases included in the sampling frame were those with financial data reported in the last three years. Within this, the ORBIS size was used if available, otherwise the imputed size. In one of the remaining three countries, Slovakia, it was necessary to use the previous year's ORBIS data as the sampling frame as the 2020 numbers of employees were heavily mis-allocated in

¹⁰ Differences of in-scope establishments between 2019 and 2020 of no more than +/- 10% were observed except in Cyprus (+33%), Slovenia (+13%) and the US (-27%). All changes observed were positive for coverage, for instance in the US overcoverage of 17% was observed in 2019 (i.e. 17% more cases on the frame than in the population), hence a reduction in numbers brought the coverage figure more in line. Similarly, Cyprus has had serious issues of under-coverage in each year, an increase in numbers closed some of this gap.

¹¹ This continuity has been maintained in all previous years with the exceptions of: Poland, where in 2018 (year three) it was necessary to use the data from the year before due to quality issues; and the same for Slovakia in the current year (2020).

ORBIS. In the remaining two countries, Cyprus and Malta, this solution resulted in severe under-coverage and too few firms to complete the survey, as seen in previous years. Here, as previously, a wider solution was chosen, of including all cases that had reported financials at any time (dropping cases with no reported financials, for which the imputation was inaccurate as it was based on the sector only). Within this, the ORBIS size was used if available and reported within the previous five years, otherwise the imputed size was used.

2.3 Sample selection and stratification

The General Module sample was selected following the steps below:

1. The number of interviews required from the fresh/cross-sectional sample was estimated at country level as the overall interview targets less the predicted number of interviews from the panel sample.
2. Target numbers of firms to select were established at the level of the interlocking size/sector cells (i.e. 16 cells per country). As noted earlier, these targets were not in proportion to the population sizes and so the sample was disproportionately stratified at this step. The US approach was the same except a boost of listed enterprises was included, by boosting the proportion of this group by a factor of two, in an attempt to generate additional numbers of these firms in the interviewed sample.
3. The selection of the target number of firms in each cell was implemented in SAS with proportional stratification by region. Variables were output according to a specification which covered the requirements for the survey (including variables needed for the interview, fieldwork implementation and monitoring, and final datasets).
4. Firms without a phone number on ORBIS were sent to Dun & Bradstreet for matching, and any that could not be matched were excluded from the fieldwork as they could not be contacted.

The cell-level targets for the sample selection were derived following an iterative process. This took account of the following:

- The target ratio of issued cases to interviews (eight cases per interview);
- The availability of phone numbers on the ORBIS sampling frame – countries with a lower proportion of populated contact details were set at a higher ratio to compensate;
- Limiting variation between weights within each size or sector category as far as possible in order to minimise the margins of error of the survey estimates. In other words, the within-sector/size cell targets were as close to the underlying population proportions as possible given the constraint of the equal-sized quotas on size/sector categories overall; and
- The numbers of available cases. In many of the countries there were limitations on the numbers of larger (and sometimes medium) firm size class cases. Where insufficient cases existed in a size class, the next-largest size class was boosted.

Following selection of an initial ‘main’ sample, to be used on commencement of fieldwork, reserve samples were selected in each country to be used for cells that were exhausted short of quota during fieldwork.

2.4 Weighting

Separate weights were required for two samples, (i) the full sample, consisting of cross-sectional cases and panel sample cases, to be used for cross-sectional analyses; and (ii) the latest wave panel sample¹² on its own, for longitudinal analyses, i.e. measurement of change between the current and previous wave.

Weighting for each group followed the same approach, of calibrating the samples to Eurostat SBS population data on the size/sector categories within each country for the EU-27 and UK, and a combination of several sources for the US (see appendix).

The weighting was conducted separately on each sample, so that each was aligned to the population totals. Basing the weighting on population figures adjusted for any differences in the covered/uncovered firm profiles in addition to making corrections to the sample where it deviated from the quota profile due to non-response. Adjusting to the total population size meant that the weights could be used for either single-country or cross-country analyses as the weighted samples reflected the correct proportions across countries.

As in previous years, the population data used was of value-added at factor cost¹³ totals. This approach gives more weight to firms with larger economic importance (based on their sector/size membership), which is a better fit to the analysis objectives of the survey than a firm-level approach which would give equal weighting to each firm in the economy. Each firm in a country/size/sector cell was given the same weight. To limit excessive damage to the sample margins of error the largest weights were capped at the 99th percentile of the distribution of weights within each country, with the exception of Ireland which was capped at the 95th percentile. This resulted in an adjustment to four countries across the full sample (Cyprus, Ireland, Luxembourg and the US), similar to the result of three countries in year four. The same capping rule was applied to the panel weights. This resulted in adjustment for 10 countries (Croatia, Cyprus, Czechia, France, Hungary, Poland, Portugal, Romania, Slovakia and the US), similar to the year four result of 12 countries.

2.4.1 Evaluation of sampling and weighting approach

This section reviews the sample quality in terms of coverage and firm-level eligibility, repeating the analysis of previous years. A consistent approach was used in all years, with ORBIS again used as the sampling frame in all countries, and the US, and also the method of case selection factoring in the imputations was the same. As in previous years, some size/sector cells fell outside of the specified coverage target range of 0.5 to 1.5 (numbers of ORBIS firms out of population totals), as shown in the table below. Compared with the previous year the trend was of no change or improved coverage in most countries, with the exception of France, where there was an increase in the level of under-coverage amongst micro and small firms. Improvements were observed in Belgium, Cyprus, Greece (although coverage remained poor), Spain, Poland and the US. All other countries had very similar levels of coverage to the previous year.

¹² Panel firms are defined as those with two consecutive waves of data.

¹³ The gross income from operating activities after adjusting for operating subsidies and indirect taxes.

Table 4: Under (pink cells)/over (orange) coverage: ORBIS sampling frame totals as a ratio of total population

	Size/ sector*	5 9	10 49	50-249	250+	Total		Size/ sector*	5 9	10 49	50-249	250+	Total
AT	M	0.51	0.73	0.91	0.96	0.58	IE	M	0.72		1.02	0.68	0.73
	S	0.35	0.56	1.07	1.06		S	0.38		0.58	0.62	0.85	
	C	0.62	0.88	1.27	1.12		C	0.66	1.62	2.31	1.06		
	I	0.39	0.71	1.14	0.92		I	0.42		0.93	1.13	0.81	
BE	M	0.61	0.93	0.90	0.96	0.8	IT	M	0.90		0.96	1.01	1.11
	S	0.58	0.99	0.98	1.11		S	1.00		1.08	1.16	1.06	
	C	0.61	1.52	0.98	1.24		C	0.81		0.94	1.12	1.32	
	I	0.58	1.27	1.03	1.09		I	0.82		1.00	1.06	1.05	
BG	M	1.21	1.14	1.02	1.04	1.33	LT	M	0.69		0.88	0.99	1.05
	S	1.29	1.38	1.26	1.32		S	0.79		0.92	1.02	1.15	
	C	1.42	1.56	1.18	0.82		C	0.67		1.02	0.90	0.94	
	I	1.69	1.53	1.35	1.33		I	1.01		1.04	1.07	1.34	
CY	M	0.01	0.03	0.27	1.83	0.46	LU	M	0.47		0.86	1.00	1.36
	S	0.02	0.05	0.14	0.62		S	0.34		0.64	0.92	1.38	
	C	0.01	0.04	0.08	1.00		C	0.49		0.77	0.89	1.06	
	I	0.02	0.03	0.13	0.14		I	0.41		0.73	0.88	0.58	
CZ	M	0.21	0.70	0.85	1.08	0.46	LV	M	0.97		1.00	0.91	0.85
	S	0.28	0.71	1.00	1.13		S	1.05		0.95	0.97	1.07	
	C	0.24	0.73	0.82	1.04		C	1.39		1.03	0.81	0.94	
	I	0.30	0.83	0.95	1.16		I	1.16		1.07	1.03	1.02	
DE	M	0.66	0.83	0.81	0.77	0.63	MT	M	0.08	0.33	0.42	0.50	
	S	0.47	0.51	0.68	0.77		S	0.06	0.29	0.35	4.00		
	C	0.68	0.81	1.01	1.01		C	0.01	0.26	0.50	4.00		
	I	0.58	0.88	0.90	0.72		I	0.07	0.35	0.60	1.92		
DK	M	0.65	0.75	0.85	0.91	0.74	NL	M	0.53		0.94	0.68	0.88
	S	0.74	0.75	0.76	0.85		S	0.62		0.92	0.70	0.69	
	C	0.66	0.90	1.06	0.81		C	0.49		1.08	0.75	0.65	
	I	0.56	0.81	0.87	0.87		I	0.56		1.08	0.74	0.61	

	Size/ sector*	5 9	10 49	50-249	250+	Total		Size/ sector*	5 9	10 49	50-249	250+	Total
EE	M	1.01	1.03	0.87	0.79	1.18	PL	M	0.50	0.39	0.69	0.87	0.67
	S	1.20	1.33	0.80	0.83		S	0.66	0.49	0.81	0.92		
	C	1.37	1.34	0.63	0.70		C	1.00	0.51	0.76	0.70		
	I	1.21	1.25	0.83	0.81		I	1.19	0.64	0.88	0.94		
EL	M	0.19	0.91	1.02	1.05	0.19	PT	M	0.73	0.94	1.06	1.20	0.91
	S	0.08	0.25	0.76	1.44		S	0.83	1.07	1.21	1.21		
	C	0.05	0.16	0.40	0.60		C	0.86	1.05	1.21	0.88		
	I	0.13	0.48	0.76	0.97		I	0.81	1.06	1.27	1.14		
ES	M	0.67	0.90	0.97	1.00	0.68	RO	M	0.80	0.85	0.86	0.91	0.88
	S	0.49	0.77	0.98	1.02		S	0.82	0.90	1.06	1.10		
	C	0.73	1.23	1.12	1.06		C	0.90	0.89	0.86	0.78		
	I	0.45	0.87	0.93	1.05		I	0.91	1.04	0.98	1.12		
FI	M	0.57	0.72	0.86	0.97	0.77	SE	M	0.61	0.90	0.92	1.01	0.82
	S	0.73	0.82	1.04	0.95		S	0.79	0.89	0.87	0.92		
	C	0.67	0.83	1.16	1.19		C	0.74	1.08	1.08	1.17		
	I	0.74	0.90	1.30	0.86		I	0.57	1.01	0.96	1.09		
FR	M	0.19	0.37	0.65	0.87	0.32	SI	M	0.84	1.13	0.99	1.04	0.99
	S	0.27	0.45	0.76	0.87		S	0.87	1.26	1.03	1.10		
	C	0.17	0.36	0.90	1.26		C	0.83	1.24	1.14	1.10		
	I	0.21	0.41	0.74	0.93		I	0.80	1.32	1.23	0.89		
HR	M	0.69	0.94	0.94	1.08	0.85	SK	M	0.36	0.74	1.01	1.00	0.76
	S	0.74	0.88	0.92	1.03		S	0.76	1.01	0.97	1.01		
	C	0.84	1.05	1.02	0.61		C	0.45	1.11	1.02	0.79		
	I	0.85	1.07	1.04	1.05		I	1.05	1.90	1.12	1.12		
HU	M	0.84	1.02	1.03	1.15	1.03	UK	M	0.87	1.05	1.08	1.42	1.22
	S	1.04	1.20	1.09	1.19		S	1.15	1.07	1.20	1.33		
	C	0.97	1.22	1.05	0.85		C	1.45	1.66	1.55	1.54		
	I	0.80	1.19	1.05	1.01		I	1.39	1.60	1.78	1.82		
							US	M	1.36	1.50	1.38	0.91	0.99
							S	0.68	0.80	0.84	0.69		

	Size/ sector*	5-9	10-49	50-249	250+	Total		Size/ sector*	5-9	10-49	50-249	250+	Total
								C	1.12	1.62	1.79	1.20	
								I	1.35	1.28	1.13	0.58	
* M = Manufacturing; S = Services; C = Construction; I = Infrastructure.													

Regarding firm-level eligibility, as reported in section 4.6.1, the eligibility rate across the EU27, UK and US was 67% for the top-up sample, similar to the previous wave of the study (which was 72%).

Finally, the sampling and weighting approach can be evaluated in terms of its ability to deliver the required margins of error, or the size of the design effect from weighting the sample¹⁴. The design effect provides a measure of the efficiency of the sample, in this instance due to weighting. A design effect of one means that the weighted sample margins of error are equal to those from a simple random sample without any weighting, i.e. as small as possible. This is a useful measure as it provides the efficiency of the sample and is independent of the overall sample size of each country.

¹⁴ The design effect from weighting is calculated using Kish's formula: the sample size multiplied by the sum of squared weights, divided by the sum of weights squared.

The table below shows the sample design effects for the four size classes and the country samples overall. For the majority of the countries, the impact from weighting on size classes was fairly modest, given the underlying sector populations were of similar sizes, meaning that aiming for equal-sized sector quotas did not deviate much from the underlying population and limited damage to precision was observed.

	<i>Micro (5-9)</i>	<i>Small (10-49)</i>	<i>Medium (50-249)</i>	<i>Large (250+)</i>	<i>Totals</i>
<i>Austria</i>	1.06	1.08	1.03	1.03	1.23
<i>Belgium</i>	1.11	1.10	1.05	1.04	1.99
<i>Bulgaria</i>	1.15	1.19	1.16	1.07	1.53
<i>Cyprus</i>	1.42	1.81	1.95	1.13	2.63
<i>Czechia</i>	1.12	1.10	1.05	1.04	1.91
<i>Germany</i>	1.17	1.15	1.14	1.04	2.21
<i>Denmark</i>	1.11	1.11	1.10	1.06	2.13
<i>Estonia</i>	1.49	1.49	1.38	1.00	4.41
<i>Greece</i>	1.28	1.29	1.14	1.13	1.40
<i>Spain</i>	1.17	1.13	1.09	1.06	1.53
<i>Finland</i>	1.11	1.08	1.08	1.05	2.15
<i>France</i>	1.11	1.12	1.09	1.08	2.41
<i>Croatia</i>	1.12	1.14	1.12	1.14	2.15
<i>Hungary</i>	1.23	1.20	1.05	1.06	3.55
<i>Ireland</i>	1.22	1.25	1.09	1.00	2.01
<i>Italy</i>	1.14	1.26	1.13	1.07	1.49
<i>Lithuania</i>	1.18	1.15	1.11	1.08	1.74
<i>Luxembourg</i>	1.21	1.27	1.20	1.18	1.66
<i>Latvia</i>	1.13	1.16	1.13	1.14	3.07
<i>Malta</i>	1.15	1.07	1.06	1.07	1.57
<i>Netherlands</i>	1.21	1.13	1.14	1.10	1.57
<i>Poland</i>	1.18	1.11	1.07	1.03	1.74
<i>Portugal</i>	1.11	1.11	1.07	1.04	1.32
<i>Romania</i>	1.13	1.15	1.14	1.06	2.38
<i>Sweden</i>	1.05	1.06	1.04	1.03	1.77
<i>Slovenia</i>	1.22	1.22	1.06	1.08	1.76
<i>Slovakia</i>	1.24	1.24	1.08	1.09	3.01
<i>United Kingdom</i>	1.16	1.19	1.23	1.16	4.34
<i>EU27 and UK</i>	2.59	2.80	3.09	3.55	6.87
<i>US</i>	1.20	1.18	1.22	1.19	3.93

The sector category design effects on the other hand tended to be a bit larger (see table below), as these were dependent on the variation in weighting of the underlying size classes, where the equal-sized quotas requirement was more out of line with the population distributions.

Table 6: Design effect from weighting, value added weight – sector categories

	<i>Manufacturing (NACE C)</i>	<i>Services (NACE G/I)</i>	<i>Construction (NACE F)</i>	<i>Infrastructure (NACE D/E/H/J)</i>	<i>Totals</i>
<i>Austria</i>	1.10	1.23	1.18	1.16	1.21
<i>Belgium</i>	1.85	1.80	1.57	1.78	1.94
<i>Bulgaria</i>	1.26	1.36	1.21	1.43	1.53
<i>Cyprus</i>	1.10	1.94	1.11	1.21	2.61
<i>Czechia</i>	1.51	1.77	1.83	1.78	1.91
<i>Germany</i>	1.69	2.05	1.59	1.80	2.18
<i>Denmark</i>	1.89	1.98	1.60	1.99	2.13
<i>Estonia</i>	4.52	6.62	6.65	1.58	4.41
<i>Greece</i>	1.18	1.16	1.08	1.16	1.38
<i>Spain</i>	1.33	1.36	1.28	1.34	1.52
<i>Finland</i>	1.77	2.00	1.84	1.97	2.11
<i>France</i>	2.03	2.35	2.21	2.19	2.39
<i>Croatia</i>	1.78	2.02	1.53	1.99	2.12
<i>Hungary</i>	2.65	2.85	2.42	2.91	3.52
<i>Ireland</i>	1.95	1.98	1.92	1.32	1.95
<i>Italy</i>	1.22	1.21	1.13	1.23	1.47
<i>Lithuania</i>	1.42	1.62	1.44	1.64	1.71
<i>Luxembourg</i>	1.65	1.27	1.64	1.53	1.65
<i>Latvia</i>	2.57	2.54	3.09	2.95	3.03
<i>Malta</i>	1.50	1.45	1.10	1.50	1.55
<i>Netherlands</i>	1.36	1.43	1.41	1.43	1.56
<i>Poland</i>	1.50	1.60	1.54	1.63	1.72
<i>Portugal</i>	1.21	1.21	1.15	1.24	1.32
<i>Romania</i>	2.02	2.27	1.99	2.16	2.37
<i>Sweden</i>	1.61	1.74	1.50	1.64	1.74
<i>Slovenia</i>	1.41	1.42	1.52	1.59	1.75
<i>Slovakia</i>	2.37	2.62	2.36	2.62	3.01
<i>United Kingdom</i>	3.40	3.70	3.50	3.62	4.30
<i>EU27 and UK</i>	5.78	6.08	5.58	6.82	6.80
<i>US</i>	3.09	3.69	3.63	3.49	3.87

3 Questionnaire translation

As in all previous waves, Ipsos undertook a rigorous translation of all module questionnaires, adhering to the renowned TRAPD model (Translation, Review, Adjudication, Pre-testing and Documentation). Using an adapted approach to this model, the English source questionnaires for all Modules underwent a rigorous seven stage translation process which were, in consecutive order: 1) review by country coordinators; 2) translation by an approved translator; 3) proof-reading by Language Connect (a specialist translation agency) translators who were independent of the initial translators; 4) review by EIB; 5) adaption where needed; 6) sign-off by EIB; and 7) script updated. The source script was scripted and signed off in English prior to the start of the translation, so the format used by translators enabled a straightforward 'overlay' of the script from English to the target language to take place. The other fieldwork materials were translated using a similarly robust, albeit slightly simplified process, which included the separate proofreading by Language Connect, but not the EIB¹⁵.

In principle, the national agencies were responsible for translating (i.e. step 2 mentioned above) all the fieldwork material and questionnaires, whereas Language Connect reviewed/proof-read translations. For six languages, however, Language Connect handled both the translation (step 2) and review (step 3), using linguists working independently of each other. This was the case for Estonian, Latvian, Lithuanian, Russian, Portuguese, Swedish, French, Slovak, Dutch, Spanish and Czech¹⁶.

For languages spoken in more than one country, the translation went through an adaptation process, meaning that one master translation was made and then adapted for local use. In this approach, an initial translation was prepared by the local agency of the country with the greatest number of native speakers of the language. Subsequently, this translation was adapted by the local agencies in countries where a local version is spoken of the same language.

All questionnaires and fieldwork materials were translated into the official languages of the countries where the survey was undertaken, insofar as deemed relevant for the purpose of this study and in agreement with Ipsos.

¹⁵ The interviewer and questionnaire manual were not proofread by Language Connect.

¹⁶ The Baltic (EE, LV, LT) country office and Portugal did not translate fieldwork materials; in these countries Language Connect carried out the translation of the questionnaires, data sheet and glossary. In Sweden, Language Connect translated the questionnaire and all fieldwork materials.

<i>Country</i>	<i>Language(s)</i>	<i>Country</i>	<i>Language(s)</i>
<i>AT</i>	German	<i>IE</i>	English
<i>BE</i>	Dutch/French	<i>IT</i>	Italian
<i>BG</i>	Bulgarian	<i>LT</i>	Lithuanian
<i>CY</i>	Greek	<i>LU</i>	French/German
<i>CZ</i>	Czech	<i>LV</i>	Latvian/Russian
<i>DE</i>	German	<i>MT</i>	Maltese/English
<i>DK</i>	Danish	<i>NL</i>	Dutch
<i>EE</i>	Estonian/Russian	<i>PL</i>	Polish
<i>EL</i>	Greek	<i>PT</i>	Portuguese
<i>ES</i>	Spanish	<i>RO</i>	Romanian
<i>FI</i>	Finnish	<i>SE</i>	Swedish
<i>FR</i>	French	<i>SI</i>	Slovenian
<i>HR</i>	Croatian	<i>SK</i>	Slovak
<i>HU</i>	Hungarian	<i>UK</i>	English (United Kingdom)
<i>US</i>	English (United States)		(Source)

4 Fieldwork

4.1 Interviewer briefings and training

Ipsos and its partner agencies used skilled interviewers with experience in working on complex business-to-business surveys, including dedicated teams with particular experience of interviewing specific types of businesses such as large businesses. In addition, extra emphasis was put on interviewer training and the development of training materials.

All country fieldwork coordinators participated in a webinar pre-fieldwork briefing session organised by Ipsos. All interviewers working on the project received bespoke training in their country office, using the training material provided by the Central Coordination Team.

4.2 Fieldwork supporting materials

In this wave, and collaborating closely again with the EIB, Ipsos further advanced the series of training materials developed in previous years to further familiarise the country fieldwork coordinators and interviewers with the project. These included: training slides, an interviewer manual, a questionnaire manual and a glossary of key terms.

The interviewer manual contained background information on the survey, best practice interviewing techniques (in general and specific to this survey), and answers to questions that might be commonly asked by respondents. The questionnaire manuals contained a summary of questionnaire-related issues with a list of the questions that might prove most problematic, and guidance for preventing any such problems occurring. The glossary of key terms allowed interviewers to familiarise themselves with technical/financial terms and served as an easy and accessible reference guide.

In addition, Ipsos again produced materials to facilitate the recruitment process which were based on the materials developed for previous waves of fieldwork. These materials consisted of an introductory letter on the survey for respondents which interviewers could email to the respondent directly from the CATI platform while they were on the phone, a dedicated survey website in all 27 EU countries, the UK and the US¹⁷, and a datasheet with key questions from the surveys which was emailed to respondents, allowing them to prepare for the interview.

The introductory letter and website provided assurance that the survey was *bona fide*, stressed the confidential nature of the study, outlined the topics covered and, ultimately, aimed to achieve high participation in the survey. For example, it stressed the importance of the study and how the findings can help businesses like theirs in the future. With an eye on inducing trust, the introductory letter and website carried both the EIB and Ipsos/local agency logos and contact details for individuals at each organisation (for the US, the introductory email included additional credentials on behalf of Columbia University). As in

¹⁷ In total, 34 specific webpages (URLs) were created when including the language specific versions in countries with more than one national language. The website was hosted on Ipsos MORI servers, see e.g. the German version: <http://eib.ipsos-mori.com/de/>

previous waves, evidence suggests that these materials, and in particular the datasheet, not only facilitated recruitment, but helped to obtain more accurate data as well.

4.3 Fieldwork platform

Computer-assisted telephone interviewing (CATI) data collection using the Dimensions platform was again used in all countries – either via full-service Dimensions software or connecting via web links. Scripting was carried out by Ipsos' central Dimensions team. Local agencies were supplied with secure access codes, and each was given a dedicated space on the Dimensions site where appropriate for them to upload their call outcome data at pre-specified intervals.

4.4 Mainstage fieldwork period

This section gives a country-by-country overview of the fieldwork progress by module and country. The information presented in the next Tables (8 and 9) generally does not include any interviews that were discarded for quality reasons. Table 8 presents the week-by-week fieldwork progress for the General Module. The table shows the cumulative number of achieved interviews in each week and the overall time it took to reach the final number of completed interviews. In several countries, fieldwork teams worked through panel members first, leading to many interviews in the first week or two of fieldwork, a slight tail-off and then a pick-up as the top-up sample was more exhaustively dialled.

Table 8: Weekly GM EIBIS fieldwork progress (completed interviews), by country

Week commencing	wc 11/05/2020	wc 18/05/2020	wc 25/05/2020	wc 01/06/2020	wc 08/06/2020	wc 15/06/2020	wc 22/06/2020	wc 29/06/2020	wc 06/07/2020	wc 13/07/2020	wc 20/07/2020	wc 27/07/2020	wc 03/08/2020	wc 10/08/2020	wc 17/08/2020
Week #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AT	134	181	235	265	295	345	409	463	480						
BE	77	94	117	150	207	246	262	294	328	384	413	448	480		
BG	89	165	208	267	309	351	389	398	416	433	451	468	480		
HR	138	197	253	307	341	386	426	473	488						
CY	42	74	90	98	105	110	119	146	166	180					
CZ	16	63	106	133	171	203	242	266	305	368	426	480	480	480	481
DK	64	105	190	221	277	296	317	337	366	366	366	366	387	432	480
EE	4	45	99	126	172	207	226	239	254	260	298	354	400		
FI	76	160	247	291	322	354	407	460	480						
FR	36	73	139	166	232	331	418	473	539	594	600	600	600	600	601
DE	139	203	283	361	417	457	486	529	567	576	584	601			
EL	58	150	325	366	395	401	403								
HU	60	145	328	360	420	438	460	475	481						
IE	69	88	123	146	163	191	228	252	291	329	346	374	401		
IT	85	176	230	274	307	348	386	417	451	489	530	551	592	592	601
LV	6	61	96	122	149	190	201	227	245	252	253	278	318	348	370
LT	3	40	136	198	245	293	351	385	394	400					
LU	39	69	90	101	118	130	130	138	145	150	157	164	176	180	
MT	17	51	79	98	116	129	146	153	159	172	180				
NL	28	48	135	171	226	274	322	357	412	460	480				
PL	79	132	190	241	265	316	347	376	408	439	469	476	478	483	483
PT	33	83	127	156	175	216	261	287	333	380	441	469	481		
RO	55	116	167	207	238	282	320	358	383	432	463	480			
SK	36	65	103	129	163	201	236	261	286	331	369	400			
SI	58	131	207	249	312	382	400								
ES	13	72	149	201	259	342	412	464	511	567	600				
SE	96	133	213	267	306	352	406	442	467	480					
UK	78	110	127	159	181	225	257	287	353	454	506	557	600	600	601
US	45	66	109	189	261	323	380	437	535	633	727	777	800		
Total	1673	3096	4901	6019	7147	8319	9347	10197	11046	11861	12451	12895	13206	13290	13372

Table 9 presents the week-by-week fieldwork progress for the Add-on Module (Municipality survey). The table depicts the number of interviews and overall timeline for reaching the final number of completed interviews.

Table 9: Weekly AOM EIBIS fieldwork progress (completed interviews), by country

<i>Week commencing</i>	wc 08/06/2020	wc 15/06/2020	wc 22/06/2020	wc 29/06/2020	wc 06/07/2020	wc 13/07/2020	wc 20/07/2020	wc 27/07/2020	wc 03/08/2020	wc 10/08/2020	wc 17/08/2020
<i>Week #</i>	1	2	3	4	5	6	7	8	9	10	11
AT	-	2	7	12	16	16	17				
BE	-	4	4	13	30	33	36	36	36	36	39
BG	2	4	6	9	10	17	18	20	27		
HR	1	6	15	18	18	18	20				
CY	-	1	1	1	3	4	4	5			
CZ	1	5	6	11	20	26	28				
DK	-	-	-	-	4	10	14	14	14	18	
EE	-	-	-	-	1	7	8	8	9		
FI	1	5	13	18							
FR	1	5	5	6	11	20	26	32	50		
DE	-	2	2	8	23	33	39	46	50		
EL	-	3	17	17	24	27	27	28			
HU	-	10	10	13	18	25	26	27			
IE	-	1	1	2	3	3	4	5	5	7	7
IT	3	21	27	29	36	43	50	50	57	57	
LV	-	-	-	-	1	1	6	8	11	14	
LT	-	-	-	-	-	13	14				
LU	-	-	1	1	1	4					
MT	-	3	3	3	5						
NL	-	1	8	12	14	16	33	40	41		
PL	-	3	13	13	13	21	36	41	41	46	50
PT	-	3	6	6	9	14	17				
RO	7	27	40	41	45						
SK	2	3	3	3	10	13	13	13	14		
SI	-	4	7								
ES	-	4	10	15	21	24	24	29	50		
SE	-	9	14	15	22	23	24	26	27		
<i>Total</i>	18	126	219	277	389	490	563	603	664	681	688

4.5 Weekly fieldwork reporting

For the duration of fieldwork, the CCT shared with EIB weekly fieldwork reports, showing the overall, country- and sector/size (or equivalent sub-group) level progress for all modules. The weekly fieldwork

reports provided, among other elements, detailed information on the sample provided, the sample dialled, remaining sample, the number of full and partially completed interviews, the number of appointments, the number of respondents who consented to being re-contacted in coming year(s), etc. In addition, the CCT produced a weekly “RAG report”. Indicating by colours (Red, Amber (yellow), and Green, i.e. “RAG”) how countries were performing, the RAG report functioned essentially as a simplified weekly report, facilitating the understanding of fieldwork progress. In addition to the weekly report and the RAG report, the CCT shared with EIB weekly fieldwork data, containing data for all completed interviews so far.

4.6 Response rates

In this section, some top-level information on response rates is provided. Across the EU27, the UK and the US, the most prevalent outcome for the General Module was refusals, which accounted for around 58,000 firms. This is considerably below the number of refusals (almost 80,000) in the first wave, slightly lower than in the fourth wave (63,000) and similar to the number in the second and third waves (58,000). The slight increase in wave 4 is driven by the US sample, as fieldwork was conducted for the first time in this market and the sample did not include panel respondents. With the panel members in the US in the fifth wave, the refusal rate is now in line with the second and third wave. There were 13,372 completed interviews, 2,063 partial interviews and 4969 ‘ineligible’ firms in this wave.

Ineligible cases included all (top-up sample) firms that failed the screener at the start of the survey – either for having fewer than 5 employees; not being a business (ever or currently); carrying out ineligible activities or being ‘out of target’. However, panel respondents remained eligible to be interviewed – even if they reported an ‘ineligible’ sector or size in the current wave of the study – unless the business had been discontinued.

A further 7,975 firms were unavailable for interview during the fieldwork period. There were a small number of countries with high numbers of respondents unavailable during fieldwork and a high number of refusals (Poland, France and the UK) as well as a small number of other countries that had a high number of refusals (the US and Latvia).

We provide the overall and by country sample outcome rates for the top-up and panel sample firms in Tables 10a-10b¹⁸. The panel sample consists of firms who participated in the previous wave of the survey, while the top-up sample refers to firms that did not participate in the previous wave.

The average eligibility rate for the top-up sample firms across the EU27, the UK and the US was 67%; the contact rate was 51%; the refusal rate was 40% and the response rate was 4%. While the response rate is broadly in line with the previous wave, eligibility, contact and refusal rates have fallen by 5%, 6%, and 5% respectively. Across the EU27, the UK and the US survey countries, the average percentage of usable numbers out of the sample dialed was 86% and is at similar levels as during the previous wave of the study (85%). The lowest percentages of usable numbers were found in Bulgaria (62%), Italy (64%), Slovakia (77%)

¹⁸ As this is a quota survey, the response rate cannot be interpreted in the same way as in a random probability survey due to the lack of a definite gross sample. Therefore, the calculated response rate is not comparable to response rates from random probability surveys. The survey outcome and response rates presented below are considered more of an indicator of the efficiency of the survey process than of data quality. Please refer to the separate contact outcomes report for call outcome definitions and outcome rate formulas used for the calculations.

and Romania (71%); the corresponding percentage is 83% in the US and it was 79% or higher in all other European countries.

As expected, the corresponding figures for panel sample firms, who were interviewed in the previous wave and consented to be re-contacted this year, are more favourable. The average eligibility rate (i.e. the proportion of firms who were still in operation) for the panel sample firms across the EU27, the UK and the US was 98%, the contact rate was 83%; and the response rate was 47%, which all are exactly the same with the previous wave. The refusal rate was 31%, which has fallen by 1%. The average percentage of usable numbers remained the same (97%).

On average, 16 eligible firms needed to be contacted to secure a top-up sample interview across the survey countries. The US had the highest number of eligible firms needed to secure an interview (55). In the UK, interviewers also needed a considerably higher number of eligible firms to secure an interview from the top-up sample (26).

Table 10a: Calculated outcome rates General Module – Top-up sample

Country	Usable numbers (out of those dialled) (%)	Ratio of usable + assumed eligible sample to interviews	Eligibility rate (%)	Contact rate (%)	Response rate (%)	Refusal rate (%)
Total (EU27 & UK & US)	86%	16.37	67%	51%	4%	40%
AT	88%	10.85	92%	43%	8%	34%
BE	88%	13.29	63%	71%	5%	59%
BG	62%	7.41	75%	40%	8%	26%
CY	85%	6.31	65%	67%	11%	51%
CZ	91%	20.01	85%	36%	4%	27%
DE	92%	16.15	86%	60%	5%	50%
DK	95%	11.43	65%	73%	6%	63%
EE	96%	9.87	70%	62%	8%	51%
EL	97%	5.42	84%	87%	16%	70%
ES	96%	11.11	68%	66%	7%	58%
FI	95%	8.52	76%	55%	10%	42%
FR	93%	17.89	87%	60%	5%	45%
HR	82%	10.75	88%	54%	7%	45%
HU	87%	12.59	82%	59%	6%	49%
IE	79%	19.57	81%	45%	4%	35%
IT	64%	16.04	74%	54%	3%	38%
LT	98%	6.61	34%	71%	9%	58%
LU	82%	11.83	85%	54%	6%	33%
LV	86%	18.64	51%	63%	3%	59%
MT	92%	2.97	36%	93%	16%	74%
NL	93%	14.76	75%	66%	5%	59%
PL	97%	12.21	37%	83%	4%	60%
PT	97%	16.78	89%	41%	6%	33%
RO	71%	13.76	55%	44%	4%	29%
SE	88%	10.55	83%	67%	8%	46%
SI	95%	5.56	84%	60%	16%	41%
SK	77%	14.43	79%	23%	5%	15%
UK	90%	26.20	48%	65%	2%	56%
US	83%	55.35	67%	23%	1%	18%

Table 10b: Calculated outcome rates General Module – Panel sample

Country	Usable numbers (out of those dialled) (%)	Ratio of usable + assumed eligible sample to interviews	Eligibility rate (%)	Contact rate (%)	Response rate (%)	Refusal rate (%)
Total (EU27 & UK & US)	97%	2.03	98%	83%	47%	31%
AT	98%	1.82	100%	92%	54%	34%
BE	96%	2.03	98%	88%	47%	35%
BG	96%	1.87	97%	78%	51%	21%
CY	93%	2.16	100%	75%	43%	27%
CZ	98%	3.92	95%	64%	25%	33%
DE	95%	1.51	98%	92%	63%	25%
DK	98%	1.80	98%	85%	54%	29%
EE	96%	2.30	95%	70%	41%	26%
EL	99%	1.63	100%	92%	61%	29%
ES	99%	2.37	99%	92%	42%	49%
FI	98%	1.87	98%	79%	52%	24%
FR	98%	2.24	98%	74%	44%	15%
HR	98%	1.77	98%	89%	55%	32%
HU	98%	3.31	97%	79%	29%	39%
IE	99%	2.48	99%	76%	40%	31%
IT	95%	1.55	99%	84%	61%	20%
LT	99%	2.09	95%	85%	46%	36%
LU	95%	1.73	97%	87%	54%	25%
LV	92%	2.34	90%	82%	38%	41%
MT	99%	1.66	95%	96%	58%	36%
NL	97%	1.58	97%	94%	61%	32%
PL	100%	2.22	95%	85%	44%	24%
PT	100%	1.96	100%	95%	51%	42%
RO	91%	2.00	97%	70%	45%	19%
SE	96%	1.84	100%	78%	52%	19%
SI	98%	1.75	95%	84%	55%	26%
SK	100%	2.66	99%	75%	37%	37%
UK	98%	3.07	96%	81%	31%	45%
US	96%	2.31	98%	77%	41%	34%

4.7 Consent to take part in future General Module survey

In total, 11,747 firms consented to be contacted in next year's survey. This represents 88% of firms that completed the General module interview in full which is an encouraging foundation to build on in the sixth wave. Finland (96%), Portugal, Italy, Greece, and Denmark (95%) recorded the highest consent to future contact rates, and Slovakia the lowest (66%).

Table 11 below displays the percentage of firms that have given their consent at the overall (EU27, UK and US) level, by country and by size of firm. The percentages show the proportion of firms consenting within each size category.

Table 11: Percentage of firms consenting to future contact – by size of firm					
Country	% of consenting firms - Total	Micro (5-9)	Small (10-49)	Medium (50-249)	Large (250+)
Total (EU27 + UK + US)	88%	85%	88%	88%	90%
AT	92%	90%	91%	94%	93%
BE	92%	90%	91%	92%	95%
BG	93%	89%	95%	92%	98%
CY	84%	83%	83%	88%	80%
CZ	73%	58%	74%	77%	80%
DE	95%	92%	94%	97%	95%
DK	92%	93%	88%	94%	93%
EE	85%	84%	86%	84%	100%
EL	95%	98%	89%	96%	97%
ES	89%	86%	87%	91%	92%
FI	96%	94%	97%	97%	98%
FR	82%	81%	80%	86%	79%
HR	92%	89%	92%	93%	96%
HU	79%	78%	84%	73%	83%
IE	91%	88%	92%	93%	100%
IT	95%	90%	96%	96%	99%
LT	87%	85%	83%	90%	94%
LU	86%	92%	89%	80%	87%
LV	88%	92%	86%	90%	86%
MT	84%	64%	87%	85%	100%
NL	72%	56%	77%	79%	68%
PL	90%	76%	87%	93%	95%
PT	95%	92%	95%	96%	99%
RO	91%	89%	94%	87%	91%
SE	86%	87%	87%	84%	87%
SI	91%	87%	90%	93%	94%
SK	66%	69%	65%	70%	54%
UK	86%	84%	87%	85%	89%
US	90%	92%	90%	85%	94%

5 Quality control and data processing

This section of the reports provides an overview of the quality assurance procedures in place for the EIBIS wave 4 survey (including information on data validation and data cleaning).

5.1 Quality control pre-fieldwork

The list below provides an overview of the steps the Central Coordination Team (CCT) and the local partners undertook before fieldwork began to maximise data quality. These include:

- Adapted TRAPD translation
- Local fieldwork coordinator training
- Interviewer training
- Pre-scripted hard and soft data logic checks
- Micro-level central script checks using dummy data

5.2 Quality control during fieldwork

This section summarises the steps the CCT and the local partners undertook *during* fieldwork to maximise data quality.

Similar to the previous waves, the Project Coordinators (PCs) in the national agencies were again responsible for conducting quality checks on the substantive computer-assisted telephone interviewing (CATI) data throughout the duration of fieldwork. All national agencies working on the General Module had extensive experience with verifying the quality of interviews, including on the 2016 and 2017, 2018 and 2019 EIBIS waves.

5.2.1 Quality checks

The quality requirements set for the national agencies remained the same as previous waves. The CCT specified that at least 10% of the completed interviews per country needed to be verified by the project team for each country. This means that 10% of interviews had to be either listened-into or back-checked. “Listening-in” refers to the QC measure in which a supervisor in a national agency listens to a live (or recorded) telephone (CATI) interview as it proceeds. “Back-checking” involves re-contacting respondents and verifying their identity/position and a number of their responses.

In order to ensure that all interviewers were covered by the quality checks, the CCT required national agencies to verify at least one interview per interviewer. Similar to the previous waves, it was deemed acceptable to have variations in the number of interviews back-checked per interviewer.

Table 12: Number of interviews and verified interviews*

	Total number of interviews completed*	Number of interviews verified (incl. both back-checks and listen-ins)	% of interviews verified (incl. both back-checks and listen-ins)
Austria	480	61	13%
Belgium	480	73	15%
Bulgaria	480	80	17%
Croatia	488	285	58%
Cyprus	180	43	24%
Czechia	481	75	16%
Denmark	480	55	11%
Estonia	400	39	10%
Finland	480	68	14%
France	601	95	16%
Germany	601	76	13%
Greece	403	130	32%
Hungary	481	54	11%
Ireland	401	44	11%
Italy	601	58	10%
Latvia	370	36	10%
Lithuania	400	53	13%
Luxembourg	180	21	12%
Malta	180	25	14%
Netherlands	480	86	18%
Poland	483	113	23%
Portugal	481	209	43%
Romania	480	102	21%
Slovakia	400	69	17%
Slovenia	400	330	83%
Spain	600	98	16%
Sweden	480	63	13%
UK	601	77	13%
USA	800	103	13%
Total	13,372	2621	20%

**All figures refer to the General Module only.*

As can be noted in Table 12, overall 20% of interviews were verified across countries (back-checks and listen-ins combined). In all previous waves, all countries verified the required 10% of interviews as a minimum. The variation between countries in the proportion of verified interviews can be explained by the methods countries use. Some national agencies listen-in by default to all (or most) interviews, while others put emphasis on a more limited number of back-checks. This does not infer any differences in overall quality.

Although listen-ins are efficient and effective, where these were not completed in real time, back-checks were used to better allow full checking of the authenticity of respondents and interviews. It was recommended to the national agencies that some quality checks in each country should be back-checks. Flexibility was allowed as some agencies already had advanced quality check procedures in place based exclusively on listen-ins.

5.2.2 Interview verification

As in previous waves, national agencies were instructed to verify, by either back-checks or listen-ins: 1) interview completion and validity; 2) respondent information and contact details; 3) screener question answers/eligibility of the business; 4) date and time of the interview; and 5) agreement on follow-up contact and sharing of data. The CCT provided questions that could be used for this purpose.

The interview verification process targeted first and foremost indicators of potentially “suspicious” interviews; random selection of interviews to be verified was only applied if there were no such indicators for the remaining cases. The national agencies were instructed to use the following quality indicators to determine whether an interview needed to be flagged as potentially suspicious:

- Interview length was less than 10 minutes or more than 40 minutes;
- Interview had a high number of “DK/Refusal” answers recorded, either in absolute terms or compared to the average number of “DK/Refusal” answers;
- Interview had a high number of “Other” answers compared to the average number of “Other” answers; and/or
- Interviewer completed a very high number of interviews, compared to the average.

The task of ensuring high-quality data was not limited to the prescribed checks described above. National agencies applied their usual quality criteria, and supervisors also conducted random checks on interviewers and interviews – for example by checking whether spontaneous answers given by the respondent were correctly assigned to the nearest pre-code (or to the “Other” code) by the interviewer. This also included checking “Other (specify)” responses entered at Q34 and Q35 in the General Module, which used a pre-coded list of banks. As a separate task, all “other” bank responses typed in by interviewers were reviewed and either assigned to an available bank code or retained as “other” as appropriate.

The described checks were complemented and supplemented by CCT data checks. For instance, the CCT checked responses entered by interviewers at numeric entry questions, such as number of employees employed three years ago, turnover, investment amounts and value of fixed assets. These responses were compared against other survey responses, size of firm and sample information, where available.

Responses were also compared against equivalent responses given in the previous wave where applicable. General data validation rules were adopted in agreement with the EIB. National agencies were asked to verify or follow up a small number of outliers and potential inconsistencies as appropriate, e.g. by listening-in, checking with interviewers or re-contacting respondents, though data was confirmed as correct in the vast majority of cases.

In Table 13 below an overview is provided of the total number of quality issues, either flagged initially by national agencies or by the CCT. This refers to *confirmed* quality issues, not to “suspicious” interviews, as not all “suspicious” interviews proved to have quality issues after verification by the national agencies. The confirmed quality issues were all caused by mistakes and misunderstandings; no cases of suspected fraud were detected.

Table 13: Number of interviews with reported quality issues*

Country	Number of quality issues	Country	Number of quality issues
Austria	0	Latvia	0
Belgium	0	Lithuania	0
Bulgaria	0	Luxembourg	0
Croatia	1	Malta	0
Cyprus	0	Netherlands	0
Czechia	11	Poland	1
Denmark	1	Portugal	0
Estonia	0	Romania	3
Finland	2	Slovakia	0
France	39	Slovenia	0
Germany	0	Spain	0
Greece	0	Sweden	0
Hungary	0	UK	0
Ireland	0	USA	0
Italy	0		

* All figures refer to the General Module.

As can be noted the number of detected quality issues was overall low although there was a higher than usual number of quality issues reported in France and Czechia, these interviews were replaced. As this was the fifth wave of the survey, supervisors knew which points required special attention when instructing the interviewers. Several countries also fed back that quality issues may have reduced due to the same interviewers being used across years.

5.3 Quality control post fieldwork

This section provides an overview of the steps the coordination team and the local partners took after fieldwork ended to maximise data quality. These include data editing and cleaning steps such as:

- Final frequency checks on all questions – to re-check routing;
- Final checks to ensure only permitted values have been inputted;
- Final checks on response distribution;
- Final checks to ensure the dataset contains no duplicate or near-duplicate records;
- Checks for either duplicate IDs or duplicate values across all variables;
- Final checks to identify any remaining impossible/implausible values;
- Final assessment of item non-response and outliers; and

- Final assessment of routing errors due to back-coding and cleaning/editing according to rules communicated by EIB.

5.3.1 Editing and data verification

(a) Data validation – Routing

Routing was specified within the CATI script and determined which respondents should answer every question (e.g. ASK ALL or ASK IF USED EXTERNAL FINANCE AT Q27).

The routing within the questionnaire was designed so that each question was only asked of a respondent if it was appropriate for them, based on their previous responses. The checks described below determine:

- where a respondent had been asked a question and should not have been; and
- where a respondent had not been asked a question but should have been.

Any issues highlighted by the syntax were investigated on a case-by-case basis. Overall the EIB survey had only very few, minor routing errors observed.

(b) Data validation – Permitted values

Every question had a list or range of permitted values. These could include permitted responses from a code list (e.g. 1 – Male, 2 – Female), a range (e.g. age is restricted to a maximum of 100), and includes 'Don't Know', 'Refused' and 'Not Applicable' (where applicable).

Every question was analysed to see the frequency of each response and any non-permitted values were highlighted. Each question was also reviewed to ensure that single or multiple responses were permitted as intended.

The permitted value checks did not yield any errors for the categorical and ordinal survey questions. Since the numeric variables such as Q10, Q14, Q44, and Q46 yielded a number of outliers in the first wave, Ipsos added "probing questions" in the questionnaire to improve data quality and avoid measurement error.

For panel firms, interviewers probed whether the respondent was certain when he/she reported an amount that was 10% below or 10% above what was reported in the last wave. Below is an example of a probing question for the question on turnover (q10). The interviewers recorded the correct amount if the respondent changed his/her response following the probe.

Q10b_label

The last time your company was interviewed, the turnover of the company was [PROG: INSERT <ANSWER PREVIOUS WAVE>]. Can you confirm this has increased to [PROG: INSERT <ANSWER CURRENT WAVE>]?

READ OUT IF NECESSARY: If not, what is the correct amount?

The last time your company was interviewed, the turnover of the company was [PROG: <ANSWER PREVIOUS WAVE>]. Can you confirm this has decreased to [PROG: INSERT <ANSWER CURRENT WAVE>]?

READ OUT IF NECESSARY: If not, what is the correct amount?

Similarly, the top-up firms were probed when they reported an amount that seemed implausible given the size of the firm. For the top-up firms, however, Ipsos fixed thresholds by firm size.

SCRIPT CHECK:

IF VALUE BETWEEN 100,000 - 999,999; PROMPT WITH WARNING MESSAGE

"Validation_Conversion_Label_1" FROM WARNING MESSAGE SECTION

IF VALUE BETWEEN 1,000,000 – 999,999,999; PROMPT WITH WARNING MESSAGE

"Validation_Conversion_Label_2" FROM WARNING MESSAGE SECTION

IF VALUE ABOVE 999,999,999; PROMPT WITH WARNING MESSAGE "Validation_Conversion_Label_3"

FROM WARNING MESSAGE SECTION

INTERVIEWER: CHECK WITH RESPONDENT ABOUT ITS ANSWER AND RECORD NEW INFORMATION IF NECESSARY. IF NO CHANGE, CLICK ON "NEXT" TO MOVE TO THE NEXT QUESTION.

Write in new number [PROG:OPEN BOX FOR NUMBER]

Do you mean [PROG: INSERT X] thousand [CURRENCY]?

Do you mean [PROG: INSERT X] million [CURRENCY]?

Do you mean [PROG: INSERT X] billion [CURRENCY]?

Overall, a small number of respondents changed their response after probing (100-200 cases), almost all of those who changed their response were panel firms. For example, 83 firms spread across the countries changed their response to the question on turnover after probing.

Furthermore, when fieldwork was completed Ipsos ran permitted values checks to flag implausible values. Particularly the outliers at Q13 – the question asking amounts invested in six different areas of investment, was followed up with Ipsos country offices, and corrections made to the data file where required. It is intended to add further validation to Q13 in the script in the next wave to further reduce or eliminate such issues.

(c) **Descriptives – Impossible and implausible responses**

Ipsos checked that the distribution of responses was logical. There were a number of objectives to this, and it meant something different for each question; for example,

- 1) Were any interviewers cheating (the distribution of responses does not make sense)?
- 2) Were any (obvious) errors made when keying in a response (an outlier that is clearly wrong)?

Q6 - The response given to the question on the number of employees (Q2) was again compared to the answer given to the question on the number of employees three years ago (Q6) during the interview. If the absolute difference between the numbers stated for the two questions differed substantially, the respondent's confirmation was requested by the interviewer.

SCRIPT CHECK: IF VALUE PROVIDED AT Q2 - CHECK Q6 VALUE WITH Q2 VALUE AND IF MORE THAN 25% DIFFERENCE

INTERVIEWER: CHECK WITH RESPONDENT ABOUT ITS ANSWER, AND RECORD NEW INFORMATION IF NECESSARY. IF NO CHANGE, CLICK ON "NEXT" TO MOVE TO THE NEXT QUESTION.

Write in new number [PROG:OPEN BOX FOR NUMBER]

Q13 – How much did your business invest in each of the following [six investment areas] with the intention of maintaining or increasing your company's future earnings? This is one the most salient questions of the survey. Some respondents answered this question in percentages rather than in numbers, while some other respondents reported very small or large numbers for certain areas of investment. Additional error messages were shown in wave 2, 3 and 4 to prevent the reporting of implausible values for this question. Although the "probing questions" helped improve the data quality, the data file still includes a small number of outliers for these questions where responses have been verified by the respondent and/or local fieldwork team.

(d) **Data validation – Duplicates and near duplicates observations**

Duplicate and near-duplicate observations can be defined as where an entry in the dataset has been duplicated (copied) or is very similar to other entries across the survey variables.

A duplicate entry in the data would mean that the responses of one participant are double counted. This would decrease the variance of survey estimates and may bias the results if there were a high number of duplicates or near-duplicates in the dataset. The objective of this analysis was to ensure that this did not happen. It was expected that if the duplicates were due to **interviewer cheating**, there would be small differences across the survey questions. These interviews are called near-duplicate interviews.

A measure of similarity was calculated to detect near-duplicate observations. This measure calculates the highest percent match between interviews across a range of the survey variables. The percent match is the ratio of the 'number of identical answers' to the 'number of variables checked' for the closest match found in data for a given interview.

In the first step, it was investigated whether the dataset included any duplicates on respondent IDs, or any duplicate interviews with identical values on all variables. The results verified that the dataset did not include any duplicates of this kind.

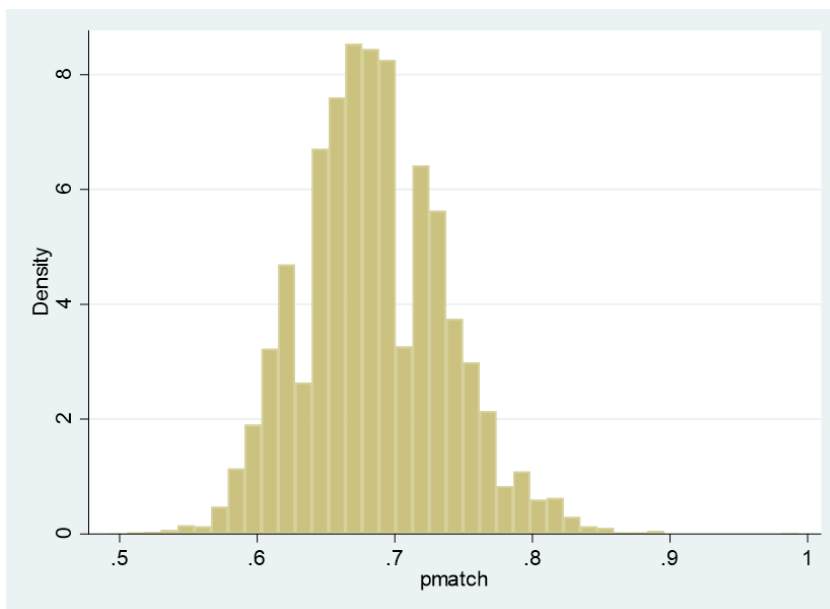
In the second step, a measure of similarity was calculated to detect near duplicate observations. This measure calculated the highest percent match between interviews across a range of the EIB investment survey variables. The percent match is the ratio of the 'number of identical answers' to the 'number of

variables checked' for the closest match found in data for a given interview.¹⁹ To this purpose, a range of substantive survey variables was used. Ipsos also excluded variables that are answered by the interviewer, and open-ended questions, but generally included as many variables as possible (the analysis covered questions q1- q61).

Near duplication can be accidental or a result of an intentional effort by interviewers in the data collection stage. If the duplicates or near duplicate interviews are due to intentional effort, the internal logic of the survey is less likely to hold. Thus, Ipsos evaluated the highly similar interviews by looking at a series of other quality indicators (e.g. level of item non-response) and case-by-case evaluation.

Overall the distribution of the percent matched did not indicate any obvious fraud (see figure 1). The distributions of percent match by country did not show any anomalies either (figure 2). Mean percentage matched was 68% and the distribution did not peak on the right end of the scale which is similar to the distributions from other surveys.²⁰

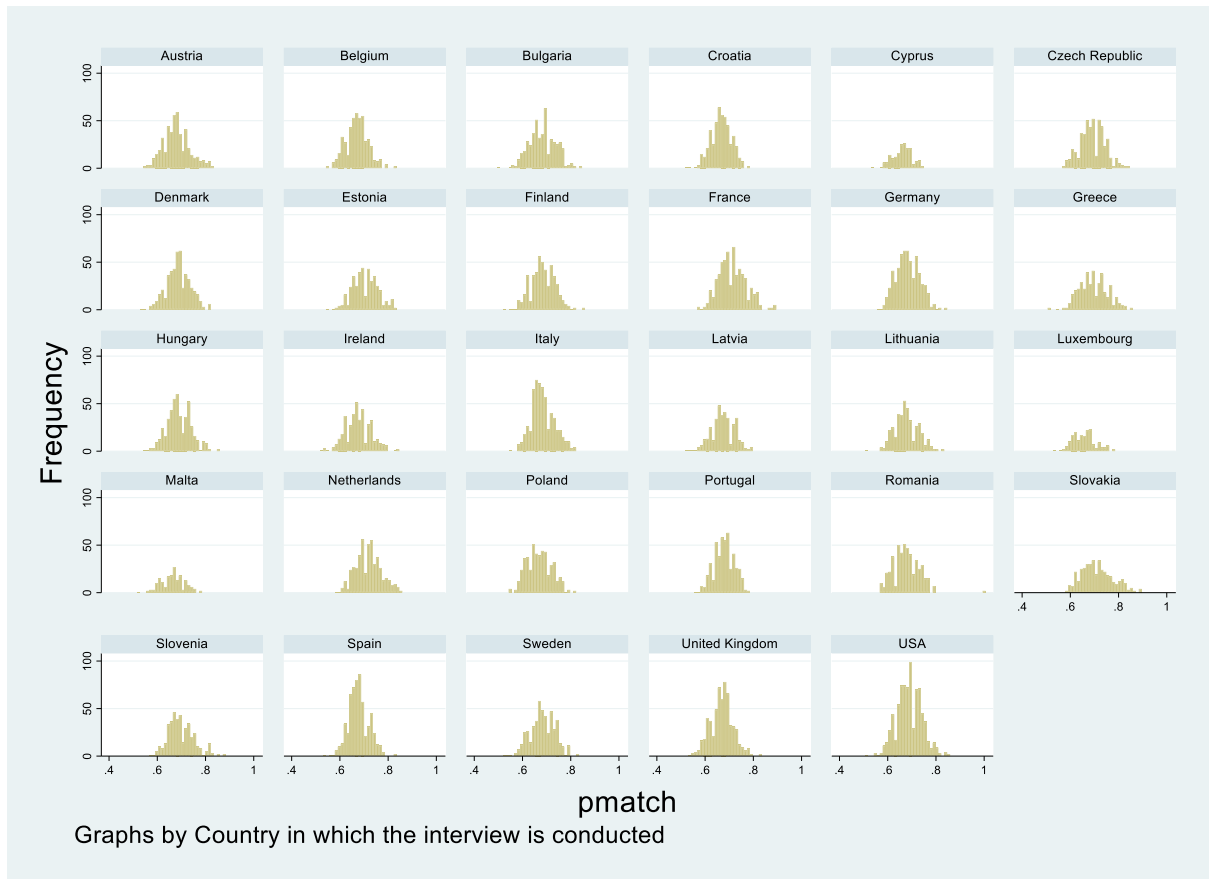
Figure 1: Distribution of percentage match – General Module



¹⁹ See PERCENTMATCH: Stata module to calculate the highest percentage match (near duplicates) between observations" by Kuriakose, N. L. (2015) for further detail. <http://econpapers.repec.org/software/bochocode/s457984.htm>

²⁰ A Kuriakose, N. L. and Robbins, M. (2015) "Falsification in Surveys: Detecting near duplicate observations" http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2580502

Figure 2: Distribution of percentage match by country – General Module



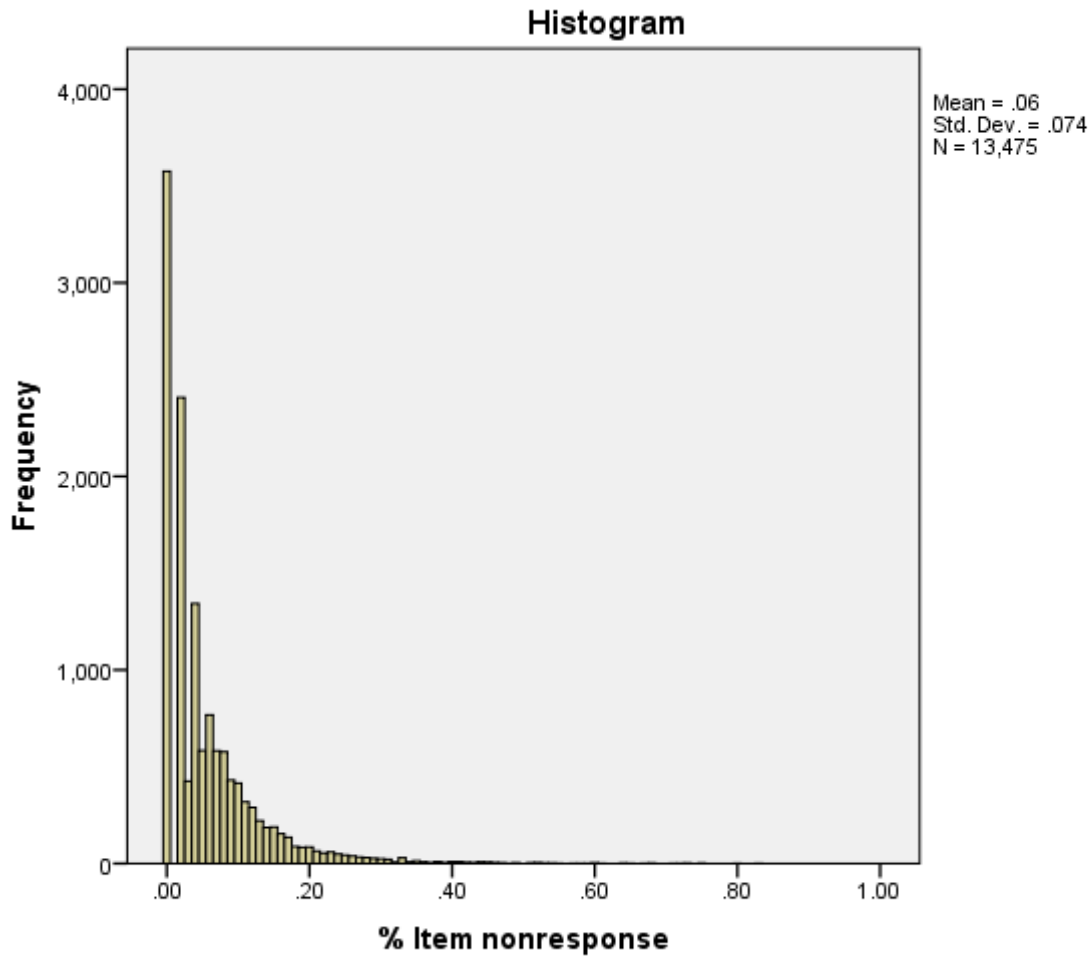
(e) Item non-response

High levels of item non-response (don't know or refused) are not desirable as they usually suggest a problem, either with the question wording or response options, or respondent disinterest or unwillingness. Analysis can be undertaken:

- 1) By respondent/firm (to see if any interviews should be removed due to high level of disinterest); and
- 2) By question (to see if there were particular questions that respondents either did not understand or did not want to answer)

On average, the proportion of item non-response by firm was 6% across all survey countries, with a standard deviation of 7.4 percentage points. Figure 3 below shows the distribution of firms with 5% or higher item non-response (firms who answered "don't know" or "refused" to more than 5% of all questions they were asked). The distribution of item non-response by firm was skewed to the right, with fewer firms having higher item non-response.

Figure 3: Item non-response in the General Module



In line with waves 1 to 4, item non-response by question was highest for numeric questions (e.g., Q44 – total fixed assets, Q46 – spend on wages), questions where respondents were asked to estimate a proportion (e.g., Q42 – proportion of commercial building stock that satisfies high efficiency standards) or approximation questions following “don’t know” answers (e.g., Q45 – approximate total fixed assets).

5.4 Back-coding

As part of the quality control, answers to semi open-ended questions were reviewed to see whether any could be back-coded into an existing pre-coded answer category. Back-coding was applied for two questions for the General Module (Q34, Q35). The recoding of these questions did not affect the routing of the other variables.

5.5 Data linking

The ORBIS dataset provides comprehensive financial and other information about the companies it covers. This information was extracted from ORBIS after fieldwork and matched with the interview IDs through the BVD IDs. Then, two separate data files including financial and patent information, keyed with the interview ID number, were prepared. Any identifying information such as contacts and personal information are excluded. These files were delivered to EIB in Stata format.

6 Appendix

Approach to derivation of US population estimates (year 5 EIBIS)

To implement the survey sampling and weighting population data were required at two levels: (i) counts of numbers of companies, and (ii) total value added; across NACE Rev. 2 sector (categories C-J) and company size (categories: 5-9, 10-49, 50-249, 250+). The data available did not fully cover these requirements and so the required values were estimated. This note describes this process.

The following data were obtained from official sources, each of which provided some of the requirements but also had some issues:

1. From the US Census Bureau data (2015) were obtained of the number of companies by detailed NAICS sector and company size category²¹. In addition to companies, this dataset contained counts of numbers of employees and total payroll, again by NAICS and company size, which was useful for some of the calculations.
 - o A problem with this data series was that the company size categories were different to those required.
2. Bureau of Economic Analysis data (2017) were obtained for value added by NAICS sector²²
 - o This data series had two issues. First, NAICS sector was available at 2-digit level only, meaning it was not sufficiently finely grained to be fully aligned with NACE on some of the sectors, affecting NACE D, E and G. Second, no breakdowns by company size could be obtained for value added.

An iterative approach was taken, using both data series. The steps are outlined below.

The first step was to re-categorise the Census Bureau data (1) into the required size categories. We fitted a Zipf distribution to describe the number of firms of size n , for $n = 0$ to 500. Based on these data number of employees is derived for each n , as $n * \text{number of companies}$.

Additionally, we fit a model for payroll per employee, using a linear fit (simple linear regression).

This was then used to split the data for the 10-99 and 100-500 groups (which each straddle a break we want, at 50 and 250) into two parts, which are then added back together into the EU size bands. This gives the required size bands for companies, and additionally for employees and payroll, which are used in some of the later steps.

The second step was to calculate value added by employee size band (2). We first dealt with the issue of sector miss-classification for US value added in NACE sectors D, E and G. We estimated value added into the NACE sectors by pro-rata-ing using total payroll (from US Census Bureau data), where NAICS can be fully aligned with NACE, the variable most closely aligned to value added. This involved identifying NAICS 2-digit sectors in the Economic Analysis value data (2) that required further breakdown to 4-digit level and applying the proportions found on payroll at the 4-digit level (US Census Bureau data, 1) to the value-added data. For reference the NAICS to NACE mapping used is below. US value added is available to 2-digit NAICS only.

²¹ <https://www2.census.gov/programs-surveys/susb/datasets/2015/>

²² https://apps.bea.gov/iTable/index_industry_gdpIndy.cfm

Table 14: NACE to NAICS mapping

NACE Description	NACE	NAICS 4-digit
Manufacturing	C	31xx-33xx
Electricity/Gas/Steam/Air conditioning	D	22xx (except 221310, 221320)
Water/Sewerage/Waste/Remediation	E	5621-5629 (plus 221310, 221320)
Construction	F	23xx
Wholesale/Retail (incl. repair of motors)	G	42xx-45xx plus 8111
Transportation/Storage	H	48xx-49xx
Accommodation/Food Service	I	72xx
Information/Communication	J	51xx

Having dealt with this we turned to the size classes. The available Bureau of Economic Analysis value data (2) had no employee size categorisation, so this step involved distributing the total value added per NAICS sector into size classes. It is known that on average a larger company gets an economy of scale and generates more per person than a smaller company, and so this distribution is not linear. We therefore assumed that value added per employee for each size band is in the same proportions as for the EU (using the EU indices). These proportions were used to “share out” US value added into the required size bands. The final results for the two key tables – company level and value added level – are shown below.

Table 15: Grouped NACE by company size – number of companies

NACE	Company size			
	5 to 9	10 to 49	50 to 249	250+
1.00 manufacturing (NACE C)	46,091	72,226	23,006	6,879
2.00 services (NACE G, I)	315,397	365,760	65,268	12,904
3.00 construction (NACE F)	105,569	97,345	14,720	2,109
4.00 infrastructure (NACE D, E, H, J)	47,329	52,665	15,833	6,453
Total	514,386	587,996	118,827	28,345

Table 16: Grouped NACE by company size –value added (millions of dollars)

NACE	Value added			
	5 to 9	10 to 49	50 to 249	250+
1.00 manufacturing (NACE C)	35,928	203,099	368,992	1,553,857
2.00 services (NACE G, I)	114,210	376,222	392,469	1,967,532
3.00 construction (NACE F)	68,975	213,797	190,946	252,085
4.00 infrastructure (NACE D, E, H, J)	30,063	113,139	187,719	1,653,261
Total	249,175	906,257	1,140,126	5,426,735

Sarah McHugh

Associate Director

sarah.mchugh@ipsos.com

Sheena Thakrar

Associate Director

Sheena.Thakrar@ipsos.com

For more information

3 Thomas More Square
London
E1W 1YW

t: +44 (0)20 3059 5000

www.ipsos-mori.com

<http://twitter.com/IpsosMORI>

About Ipsos MORI's Social Research Institute

The Social Research Institute works closely with national governments, local public services and the not-for-profit sector. Its c.200 research staff focus on public service and policy issues. Each has expertise in a particular part of the public sector, ensuring we have a detailed understanding of specific sectors and policy challenges. This, combined with our methods and communications expertise, helps ensure that our research makes a difference for decision makers and communities.