

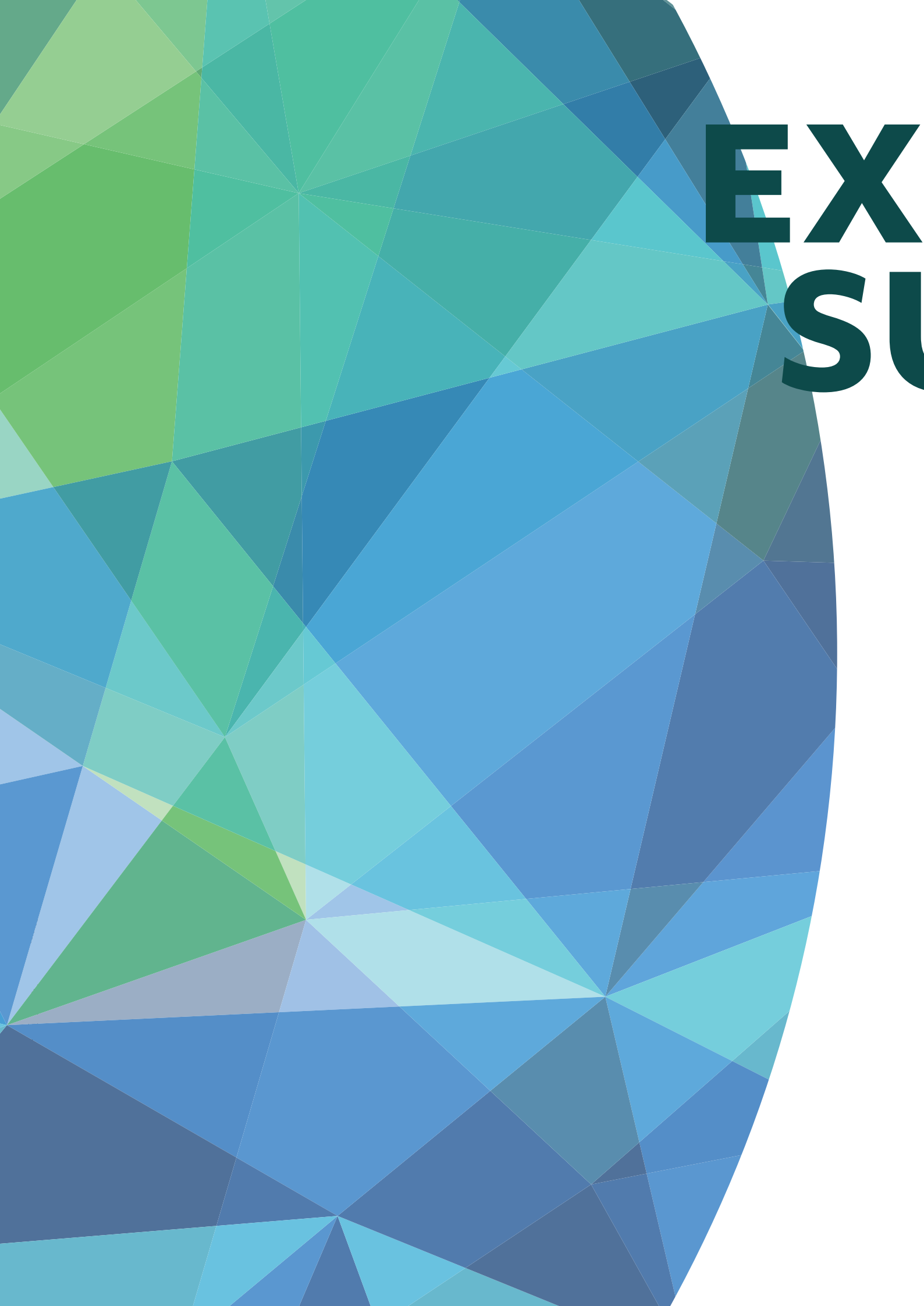
DIGITAL TRANSFORMATION IN BUSINESS

— METHODOLOGY APPENDIX

2020

Copenhagen
Economics

CE



EX
SU

EXECUTIVE SUMMARY

The Facebook company commissioned this study to measure the economic benefits of its apps and technologies for businesses.¹ 'Facebook apps and technologies' include Facebook, Instagram, Messenger, and WhatsApp.² Notably, this study includes both businesses that used Facebook apps and technologies as free-to-use tools and also those that engaged with paid advertising. The economic benefits are measured in terms of gross value added (GVA)³ and exports that businesses have credited to their use of Facebook apps and technologies. This paper covers the methodology applied for this study.

The estimates derived from the analysis should be viewed as gross figures; they do not take into account any substitution effects, displacement effects, or cannibalisation that may occur as a result of the use of Facebook apps and technologies or their potential substitutes. In other words, the study provides a snapshot of the aggregate business activity that currently leverages Facebook apps and technologies. The inherent challenges of defining and modelling a counterfactual world without online activity, or without online tools such as Facebook apps and technologies, are beyond the scope of this study.

At the time of writing, there is no single accepted methodology for estimating the impact of digital platforms on multi-sided markets. Other public studies have used willingness-to-pay formulations and return-on-investments estimates for digital advertising, amongst other techniques. As such, guidance from the literature is sparse, and methodologies used by previous studies either use internal data⁴ or are presented with a series of limitations that should be acknowledged at the outset. In this study, similar caveats apply and are discussed in detail throughout this Methodology Appendix.

Throughout this study, the research team did not have access to any internal Facebook company data regarding business use of Facebook apps and technologies, nor the value derived from their use. The methodology included a global survey supported and augmented by current and available public data with regards to employment counts, output, and GVA. In reference to the above points, these estimates should be viewed as approximate calculations using the available data.

¹ This study captured the impact on businesses' total sales, not total revenues, noting that revenues could arise from sources on which the Facebook company may have no impact.

² The term 'Facebook' is used when referencing the app. Instances of 'Facebook company' indicate the corporate entity, not the app.

³ The component of gross domestic product that captures the contribution by industry.

⁴ See Google (2018). *Economic Impact Report - United States*.

The research identified an increase in sales that businesses credited to Facebook apps and technologies, and the corresponding economic values associated with the following perspectives:

- 01** Economic activity – GVA supported by the increase in sales credited to the Facebook company
- 02** Jobs – employment supported by the increase in sales credited to the Facebook company
- 03** Exports – the increase in international sales credited to the Facebook company
- 04** Financial inclusion – supporting underrepresented or disadvantaged people to start and grow their businesses

The analysis covered private businesses in the 'market economy'. Public sector industries are not included as their revenues are most likely to be attributed to taxpayers.

The aggregate economic contribution of the Facebook company was estimated using:

- A survey of 15,342 businesses from 30 countries and regions⁵ to identify insights on their use (or non-use) of Facebook apps and technologies for business.⁶
- A collation of publicly available economic data by industry across every country and region within the survey sample.

Survey

The survey was designed by the research team and was fielded by Qualtrics between 1 October and 1 November 2019.⁷ Sampling quotas were set based on the distribution of GVA across industries for each country or region to ensure appropriate coverage of the sources of GVA within each economy. The survey was pre-tested on approximately 5 to 10 per cent of respondents from each country or region to ensure the survey was operating as intended. The survey used a two-step translation using native speakers for the non-English language survey.

⁵ The countries and regions surveyed were: Argentina, Australia, Bangladesh, Brazil, Canada, Chile, Colombia, Egypt, France, Germany, India, Indonesia, Italy, Japan, Malaysia, Mexico, Netherlands, Nigeria, Pakistan, Peru, Philippines, Poland, South Africa, Spain, Taiwan (China), Thailand, Turkey, United Kingdom, United States, and Vietnam. Approximately 500 employees were surveyed in each of the 30 countries and regions. Note that the terms 'country' and 'region', used interchangeably with 'economy' in this report, do not imply political independence but refers to any territory for which authorities report separate social or economic statistics.

⁶ The total GVA of the 30 surveyed countries and regions was 66.5 per cent of global GVA in 2015. World Bank (2018). Gross value added at basic prices (GVA) (current USD). Available at: <https://data.worldbank.org/indicator/NY.GDP.FCST.CD>. Accessed September 2019. The total GVA for the 30 countries and regions selected was \$43.26 trillion (USD) in 2015, and the total GVA for all countries and regions was \$65.05 trillion (USD) in 2015.

⁷ This was an opt-in survey that offered a small incentive of \$5-10 (USD). Employees were used as proxy informants of the business they represented. This is inferior to auditable and reliable information regarding a business. Reverse-worded questions were used as well as positively phrased questions regarding the sales attribution. The negatively phrased questions were used within the economic analysis as this produced a more conservative response (approximately 5 per cent lower attribution of sales to their use).

Screening questions were used to ensure that each respondent was currently employed and had input into the decision-making process for the business they represented. This helped to ensure respondents would have access to sufficiently detailed business information required to accurately complete the survey.

To further ensure the validity of survey data used in the analysis, incorrect, inaccurately formatted, or repeated data was removed from the sample, and a rule-based approach to removing non-feasible responses was applied.⁸

The survey was administered via online panels. These panels are unlikely to represent the true distribution of businesses by key geodemographic variables (e.g. size, industry, country/region), introducing systematic error into sample estimates. As the distribution of relevant geodemographic variables in the population is not known in all countries and regions surveyed, it is not possible to correct for this bias in a valid and reliable way, e.g. by post-weighting to match known population characteristics.

Efforts were made to ensure the survey sample was representative of the target population through quota sampling. A sampling plan was used that closely aligned to the distribution of GVA by industry for each country/region as this is a known population characteristic. However, as quota sampling is a non-probabilistic sampling technique, potential remains for selection bias. Further information can be found within this complete Methodology Appendix.

Economic analysis

Survey results and public data were used to estimate the economic contribution supported by the Facebook company in terms of GVA and exports according to the procedures as outlined in Table 1 below.

Table 1: Description of Equation Variables

Symbol	Description
c	Country/Region
i	Industry
b	Business
%Sales	Lower bound of the proportion of a respondent's sales attributed to the Facebook company according to the survey
\$Sales	A respondent's total sales according to the survey
GVA (<i>public</i>)	Total gross value added according to public data
Output (<i>public</i>)	Total output according to public data
GVA	The estimated GVA attributed to the Facebook company
Employment (<i>public</i>)	Total number of employees in an industry according to public data
%IntlSales	Lower bound of the proportion of a respondent's international sales attributed to the Facebook company according to the survey
\$IntlSales	A respondent's international sales according to the survey

⁸ See Section 3.2.5 for a detailed overview of survey scrubbing and detection of outliers.

GROSS VALUE ADDED (GVA)

To obtain the contribution to GVA from businesses' use of Facebook apps and technologies, the following method was used:

- Calculate the dollar figure of the sales attributable to the Facebook company for our sample⁹ by industry and country/region.¹⁰
- Multiply in-sample sales attributable to the Facebook company by the industry and country/region-specific ratio of GVA to output according to publicly available sources.¹¹
- Scale up the in-sample GVA attributable to the Facebook company to the industry level in each country/region, using the ratio of industry-level GVA (from public data) to the in-sample GVA reported for each industry.
- Sum GVA across industries and countries/regions for an overall estimate of GVA supported by the Facebook company for the 'market economy' in the 30 countries and regions surveyed.

This equation is conceptualised in Figure 1.

Figure 1: Calculation of GVA Attributed to Business Use of Facebook Apps and Technologies

$$GVA = \sum_{c,i} \left(\sum_b \left[\left(\frac{GVA_{c,i}(\text{public})}{Output_{c,i}(\text{public})} \right) * (\%Sales_{c,i,b}) * (\$Sales_{c,i,b}) \right] * \frac{GVA_{c,i}(\text{public})}{\sum_b \$Sales * \frac{GVA_{c,i}(\text{public})}{Output_{c,i}(\text{public})}} \right)$$

This equation can be simplified as seen in Figure 2.

Figure 2: Simplified Calculation of GVA Attributed to Business Use of Facebook Apps and Technologies

$$GVA = \sum_{c,i} \left(\sum_b [\%Sales_{c,i,b} * \$Sales_{c,i,b}] * \frac{GVA_{c,i}(\text{public})}{\sum_b \$Sales} \right)$$

⁹ 'Attributable to the Facebook company' is defined within this study as survey respondents reporting the value of sales that occur with the Facebook company's assistance. It draws from the key question of 'In the past 12 months, how much do you think your total sales would have decreased if you did not use Facebook apps and technologies?' Conceptually, this is a hard question for respondents to answer because it requires imagining a counterfactual world without the Facebook company. Therefore, the estimates should be viewed as a representation of the survey findings, applied to the national level across the 30 countries and regions and could include inaccuracies.

¹⁰ As respondents were asked for the proportion of sales they credited to the Facebook company from both a positively and negatively framed question. Positively framed question: In the past 12 months, what proportion (%) of total sales in your business relied on Facebook apps and

technologies? For example, this could be sales from targeted advertising or through people finding your business through Facebook apps and technologies. Negatively framed question: In the past 12 months, how much do you think your total sales would have decreased if you did not use Facebook apps and technologies? The responses to the negatively framed question were marginally lower than the positively framed question; this framing was chosen to be used in the study and thus produced a more conservative estimate. In answering these questions, respondents were provided percentage intervals, i.e. 11 to 20 per cent. The lower bound of this range was chosen as the most conservative measure to avoid overestimating the impact of the Facebook company.

¹¹ Public data was collated for: employment by industry, GVA by industry, output by industry, and exports of goods and services.

EXPORTS

The steps to calculate exports are:

- Calculate international sales attributable to Facebook apps and technology for our sample¹² by industry and country/region.
- Scale up the in-sample international sales attributable to the Facebook company to the industry level in each country/region, using the ratio of industry-level GVA (from public data), to the in-sample GVA reported for each industry.
- Sum the international sales attributable to the Facebook company across industries and economies, for an overall estimate across the 30 market economies surveyed.

This equation is shown in Figure 3.

Figure 3: Calculation of Exports Attributed to Business Use of Facebook Apps and Technologies

$$Exports = \sum_{c,i} \left(\sum_b [(\%IntlSales_{c,i,b}) * (\$IntlSales_{c,i,b})] * \frac{GVA_{c,i} (public)}{\sum_b \$Sales * \frac{GVA_{c,i} (public)}{Output_{c,i} (public)}} \right)$$

The estimates of the Facebook company's contribution to GVA and exports, calculated using the equations presented above, were corrected for internet penetration, controlling for selection biases inherent to using an online survey to obtain business data.¹³

¹² International sales attributable to the Facebook company are defined within this study as survey respondents reporting the value of international sales (i.e. foreign sales from the domestic country/region) which they credit to the Facebook company.

¹³ See Section 2.3 for more information. As members of online survey panels, by definition, have access to the internet, sampling strategies supported by online panels are likely to overrepresent businesses that are more likely to be on-platform.

3

EXECUTIVE SUMMARY

11

CHAPTER I
BACKGROUND AND OVERVIEW

13

CHAPTER II
ANALYTICAL METHOD

18

CHAPTER III
DATA COLLECTION

33

APPENDIX

Term / Abbreviation	Definition
ABS	Australian Bureau of Statistics
CLT	Central Limit Theorem
DOSM	Department of Statistics, Malaysia
The Facebook company	Facebook apps and technologies (including: Facebook, Instagram, Messenger, and WhatsApp)
FX	Foreign Exchange
GDP	Gross Domestic Product
GVA	Gross Value Added
Heckman Correction	The Heckman two-step correction is a statistical technique to correct bias from non-randomly selected samples or otherwise incidentally truncated dependent variables
ILO	International Labour Organization
IMF	International Monetary Fund
ISIC	International Standard Industrial Classification
Market economy	The whole economy minus industries dominated by the public sector (public administration and defence, compulsory social security, education, human health, and social work activities)
OECD	Organisation for Economic Co-operation and Development
Off-platform	Does not use any of the Facebook apps or technologies
On-platform	Uses at least one of the Facebook apps or technologies
PPP	Purchasing Power Parity
R-Squared	R-squared is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable
RBA	Reserve Bank of Australia
REV4	Revision 4
ROI	Return on Investment



METHO AI

DOLOGY PPENDIX

1. BACKGROUND AND OVERVIEW

The Facebook company commissioned this study to measure the economic benefits of its apps and technologies for businesses.¹⁴ 'Facebook apps and technologies' include Facebook, Instagram, Messenger, and WhatsApp.¹⁵ Notably, this study includes both businesses that used Facebook apps and technologies as free-to-use tools and also those that engaged with paid advertising. The economic benefits are measured in terms of gross value added (GVA)¹⁶ and exports, that businesses have credited to their use of Facebook apps and technologies. This paper covers the methodology applied to the present study.

The estimates derived from the analysis should be viewed as gross figures; they do not take into account any substitution effects, displacement effects, or cannibalisation that may occur as a result of the use of Facebook apps and technologies or their potential substitutes. In other words, the study provides a snapshot of the aggregate business activity that currently leverages Facebook apps and technologies. The inherent challenges of defining and modelling a world without online activity, or without online tools such as Facebook apps and technologies, are beyond the scope of this study.

At the time of writing, there is no single accepted methodology for estimating the impact of digital platforms and multi-sided markets. Other public studies have used willingness-to-pay formulations and return-on-investment estimates for digital advertising, amongst other techniques. As such, guidance from the literature is sparse, and methodologies used by previous studies either use internal data¹⁷ or are presented with a series of limitations that should be acknowledged at the outset.

In this study, similar caveats apply and are discussed in detail throughout this Methodology Appendix.

¹⁴ This study captured the impact on businesses' total sales, not total revenues, noting that revenues could arise from sources on which the Facebook company may have no impact.

¹⁵ The term 'Facebook' is used when referencing the app. Instances of 'Facebook company' indicate the corporate entity, not the app.

¹⁶ The component of gross domestic product that captures the contribution by industry.

¹⁷ See Google (2018). *Economic Impact Report - United States*.

1.1 OVERVIEW OF APPROACH DEVELOPMENT

Throughout this study, the research team did not have access to any internal Facebook company data regarding businesses' use of Facebook apps and technologies, nor the value derived from their use. The methodology included a global survey supported and augmented by current and available public data regarding employment counts, output, and GVA. In reference to the above points, these estimates should be viewed as an approximate calculation using the available data.

The research identified an increase in sales that businesses credited to Facebook apps and technologies, and the corresponding economic values associated with the following perspectives:

- 01** Economic activity – GVA supported by the increase in sales credited to the Facebook company
- 02** Jobs – employment supported by the increase in sales credited to the Facebook company
- 03** Exports – the increase in international sales credited to the Facebook company
- 04** Financial inclusion – supporting underrepresented or disadvantaged people to start and grow their businesses

The analysis covered private businesses in the 'market economy'. Public sector industries are not included as their revenues are most likely to be attributed to taxpayers.

The aggregate economic contribution of the Facebook company was estimated using:

- A survey of 15,342 businesses from 30 countries and regions¹⁸ to identify insights on their use (or non-use) of Facebook apps and technologies for business.¹⁹
- A collation of publicly available economic data by industry across every country and region within the survey sample.

The remainder of this document presents the analytical method to derive the economic estimates and a detailed description of the survey design.

¹⁸ Further information on the countries and regions included within this analysis can be found in Section 3.2.1.

¹⁹ The total GVA of the 30 surveyed countries and regions was 66.5 per cent of global GVA in 2015. World Bank (2018). Gross value added at basic prices (GVA) (current USD). Available at: <https://data.worldbank.org/indicator/NY.GDP.FCST.CD>. Accessed September 2019. The total GVA for the 30 countries and regions selected was \$43.26 trillion (USD) in 2015, and the total GVA for all countries and regions was \$65.05 trillion (USD) in 2015.

2. ANALYTICAL METHOD

In this section we detail the calculations used to generate the economic measures.

The calculations are undertaken at an industry level and then summed for the market economy across all surveyed countries and regions. The equations within this section use variable notation consistent with Table 2.

Table 2: Description of Equation Variables

Symbol	Description
c	Country/Region
i	Industry
b	Business
%Sales	Lower bound of the proportion of a respondent's sales attributed to the Facebook company according to the survey
\$Sales	A respondent's total sales according to the survey
GVA (<i>public</i>)	Total gross value added according to public data
Output (<i>public</i>)	Total output according to public data
GVA	The estimated GVA attributed to the Facebook company
Employment (<i>public</i>)	Total number of employees in an industry according to public data
%IntlSales	Lower bound of the proportion of a respondent's international sales attributed to the Facebook company according to the survey
\$IntlSales	A respondent's international sales according to the survey

13

2.1 GROSS VALUE ADDED (GVA)

Calculating the contribution to GVA from businesses' use of Facebook apps and technologies used the following method:

- Calculate the dollar figure of the sales attributable to the Facebook company for our sample²⁰ by industry and country/region.²¹
- Multiply in-sample sales attributable to the Facebook company by the industry and country/region-specific ratio of GVA to output according to publicly available sources.²²

²⁰ 'Attributable to Facebook' is defined within this study as survey respondents reporting the value of sales that occur with the Facebook company's assistance. It draws from the key question of 'In the past 12 months, how much do you think your total sales would have decreased if you did not use Facebook apps and technologies?' Conceptually, this is a hard question for respondents to answer because it requires imagining a counterfactual world without the Facebook company. Therefore, the estimates should be viewed as a representation of the survey findings, applied to the national level across the 30 countries and regions and could include inaccuracies. Further information can be found in Section 2.3.

²¹ As respondents were asked for the proportion of sales they credited to the Facebook company from both a positively and negatively framed questions, i.e. Positively framed question: In the past 12 months, what proportion (%) of total sales in your business

relied on Facebook apps and technologies? For example, this could be sales from targeted advertising or through people finding your business through Facebook apps and technologies. Negatively framed question: In the past 12 months, how much do you think your total sales would have decreased if you did not use Facebook apps and technologies? The responses to the negatively framed question were marginally lower than the positively framed question; this framing was chosen to be used in the study and thus produced a more conservative estimate. In answering these questions, respondents were provided percentage intervals, i.e. 11 to 20 per cent. The lower bound of this range was chosen as the most conservative measure to avoid overestimating the impact of the Facebook company.

²² Public data was collated for: employment by industry, GVA by industry, output by industry, and exports of goods and services.

- Scale up the in-sample GVA attributable to the Facebook company to the industry level in each country/region, using the ratio of industry level GVA (from public data) to the in-sample GVA reported for each industry.
- Sum GVA across industries and countries/regions for an overall estimate of GVA supported by the Facebook company for the 'market economy' in the 30 countries and regions surveyed.

This equation is conceptualised in Figure 1.

Figure 1: Calculation of GVA Attributed to Business Use of Facebook Apps and Technologies

$$GVA = \sum_{c,i} \left(\sum_b \left[\left(\frac{GVA_{c,i}(\text{public})}{Output_{c,i}(\text{public})} \right) * (\%Sales_{c,i,b}) * (\$Sales_{c,i,b}) \right] * \frac{GVA_{c,i}(\text{public})}{\sum_b \$Sales * \frac{GVA_{c,i}(\text{public})}{Output_{c,i}(\text{public})}} \right)$$

This equation can be simplified as seen in Figure 2.

Figure 2: Simplified Calculation of GVA Attributed to Business Use of Facebook Apps and Technologies

$$GVA = \sum_{c,i} \left(\sum_b [\%Sales_{c,i,b} * \$Sales_{c,i,b}] * \frac{GVA_{c,i}(\text{public})}{\sum_b \$Sales} \right)$$

2.2 EXPORTS

The steps to calculate exports are:

- Calculate international sales attributed to the Facebook company for our sample²³ by industry and country/region.
- Scale up the in-sample international sales attributed to the Facebook company to the industry level in each country/region, using the ratio of industry level GVA (from public data) to the in-sample GVA reported for each industry.
- Sum the international sales attributed to the Facebook company across industries and economies for an overall estimate across the 30 market economies surveyed.

This equation is shown in Figure 3.

Figure 3: Calculation of Exports Attributed to Business Use of Facebook Apps and Technologies

$$Exports = \sum_{c,i} \left(\sum_b [(\%IntlSales_{c,i,b}) * (\$IntlSales_{c,i,b})] * \frac{GVA_{c,i}(\text{public})}{\sum_b \$Sales * \frac{GVA_{c,i}(\text{public})}{Output_{c,i}(\text{public})}} \right)$$

²³ International sales credited to the Facebook company are defined within this study as survey respondents reporting the value of international sales (i.e. foreign sales from the domestic country/region) which they credit to the Facebook company.

2.3 OTHER CALCULATION CONSIDERATIONS

The following points capture refinements, analysis, and other aspects of the methodology that are important to note.

Quantifying the proportion of sales credited to the Facebook company

Survey questions were used to capture increases in total sales and international sales that the respondent credited to the business use of Facebook apps and technologies. This question is cognitively difficult to answer accurately, given other similar platforms could be used to derive sales in a manner similar to Facebook apps and technologies. The vast majority of businesses use multi-channel marketing approaches and therefore attempting to dissect this can lead to erroneous results.

As businesses are able to access engagement rates and other statistics through their use of Facebook apps and technologies, it is reasonable to suggest that users have a relatively informed judgement on how impactful Facebook apps and technologies are on their business' total sales, compared to other channels.

In an effort to limit any overestimation through survey questions, the respondent was asked a selection of questions regarding the sales the respondent perceives to arise from their business' use of Facebook apps and technologies. The question which produced the most conservative response was used within the economic analysis, as well as taking the lower bound of the selected answer (i.e. respondents were asked to choose the most suitable answer from ranges of 10 percentage points). Furthermore, responses where respondents selected they 'did not know' whether their business' use of Facebook apps and technologies had impacted sales were assigned a value of zero (i.e. their use of Facebook apps and technologies had not led to an increase in sales).

This method was chosen as the research team did not have access to internal Facebook company data that could provide statistics on advertising spend, click-through rates, and engagement statistics, amongst other useful data points. The responses to the questions regarding sales attribution to Facebook apps and technologies were tested for their distribution and produced low coefficients of variation.

Insufficient sample sizes to be able to calculate economic measures at the industry level for some industries

The economic analysis is undertaken at the industry level for each of the surveyed countries and regions. In some instances there were a low number of responses at the industry level and estimating the total industry's economic impacts off a small sample size would have increased the calculation error (see Table C in the Appendix A.2). To minimise the noise from using a small sample that is less representative of the population, a refinement to the methodology was required.

The target sample size for the online survey was determined by balancing the study's objectives and constraints (see Section 3.2.1 for a description of the sample size design). The target sample size meant we were unable to obtain in all occasions a sufficiently large sample for each industry across all countries and regions to estimate the Facebook company's contribution to economic measures with a consistent level of precision.²⁴ In an effort to reduce potential bias and thus maximise the precision of these estimates, in those specific occasions where this limitation emerged, we aggregated responses for a given industry across a group of similar countries and regions.²⁵ This is based on businesses within a given industry being more homogenous (including the usage of Facebook apps and technologies) than businesses within a given country or region. When sample sizes for any industry in any country or region were below 20 (i.e. in 96 out of 330 sectors across the 30 countries and regions), all values used in calculating the economic measures (e.g. total sales, international sales, and the proportion of sales supported by the Facebook company) were imputed. The imputation used the sample estimate for the indicated industry, in countries and regions within the same income group, as defined by the World Bank,²⁶ and weighted this according to the size of the industry within that country or region.

Aggregation of countries and regions into income groups for countries and regions with insufficient sample sizes reduces the impact of random error on economic estimates, increasing the validity of overall results. However, this correction reduces specificity and is dependent on the assumption that the aggregated countries and regions are homogenous (aggregation homogeneity), with each individual country or region being representative of the group. As stated above, in a best attempt to meet this assumption, countries and regions were grouped based on the World Bank's defined income groups.²⁷ The net impact of this imputation was a reduction in estimated GVA supported by the Facebook company of approximately 8 per cent.

Correcting for differences between the survey focus and the economy-wide activity

The survey was designed to estimate population characteristics relevant to an assessment of economic value via the effect of Facebook apps and technologies on business sales. However, not all organisations are oriented along sales principles, e.g. (most of) the public sector. As such, the Facebook company's impact was assumed to

²⁴ While the central limit theorem (CLT) is frequently acknowledged as holding true with sample sizes over 30, financial metrics collected through the online survey (total sales and international sales) are positively skewed, with a large degree of variation. Where distributions are skewed, we anticipate that the CLT may take more than $n \sim 100$ to be an accurate approximation of the finite sample.

²⁵ As defined by the World Bank's income groups.

²⁶ For example, our sample includes eight responses from agriculture, hunting, forestry, and fishing in Argentina; rather than estimating the economic contribution of the Facebook company for this industry in Argentina by scaling up the summed value reported by these eight responses, economic

calculations for this sector in Argentina were carried out using sample estimates for agriculture, hunting, forestry, and fishing businesses in upper-middle income countries and regions.

²⁷ The World Bank classifies the world's economies into four income groups – high, upper-middle, lower-middle, and low. This assignment is based on Gross National Income per capita (current USD) calculated using the Atlas method. The classification is updated each year on 1 July. World Bank (2019), World Bank Country and Lending Groups, Country classification, Available at: <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups> Accessed October 2019.

be zero for businesses in the following three industries, which consist predominantly of public sector organisations:

- Public administration and defence; compulsory social security
- Education
- Human health and social work activities

Through excluding these three industries, the estimates calculated through the methodology described within this paper do not assume any benefits to these industries from their use of Facebook apps and technologies. This is a more conservative approach than attempting to calculate the benefits from the public sectors' use of Facebook apps and technologies.

Correcting for differences in internet penetration

The survey was administered via online panels. Thus, by design, the survey cannot capture individuals that do not use the internet and which, by definition, cannot use Facebook apps and technologies as business tools. Given the breadth of the global study, this is considered more of an issue for developing countries and regions within our sample, where internet penetration is typically lower. As estimates of the Facebook company's economic contribution are calculated using sample estimates, this bias towards online businesses may contribute to an overestimation of the value supported by the Facebook company. For this reason, we explored possible ways to control, insofar as possible, the inherent selection bias (see Section 3.2.4 – Accounting for bias, for further information).

Internet penetration rates for each country and region were obtained from International Telecommunication Union Statistics.^{28, 29} In the calculation of the final GVA and exports figures, the internet penetration rates by country/region are multiplied by the scaled GVA and exports estimated to be credited to the Facebook company.

Measurement error

The use of survey data as an input to economic calculations introduces measurement error to the estimates provided relating to the Facebook company's economic impact. When used in calculations, due to error propagation, the magnitude of measurement error increases, decreasing the precision of calculation results. In order to gauge the overall estimate sensitivity, the skewness and precision of component variables were

²⁸ International Telecommunication Union Statistics (2017). Percentage of Individuals using the Internet. Available at: <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> Accessed October 2019.

²⁹ ITU statistics refer to internet penetration for consumers rather than businesses. As no reliable and comparable source of business internet penetration exists for all 30 countries and regions in the research scope, this was used as the best available source to apply the adjustment discussed above. Note that the use of consumer data will produce a more conservative result, which is preferred to overestimating the value supported by Facebook apps and technologies.

investigated. The key variables included in the calculations can be considered in two groups:

- Financial information, provided as a numeric response.
- Proportion of sales supported by Facebook apps and technologies, provided as a range, of which we have taken the lower bound.

When analysed for kurtosis, skewness, and precision, the first of these groups is positively skewed, with a large degree of variation. As a result, confidence intervals contain zero, with imprecise estimates. In comparison, while slightly skewed, responses to the second of these groups (proportion), the data is cleaner, with a smaller coefficient of variation, contributing to more precise estimates.

We note that the observed skew and substantial variance on key variables are likely to impact the precision of economic calculations and recommend due caution in how results are communicated.

3. DATA COLLECTION

The chosen methodology and calculations described in the previous section are based on inputs from both public and survey data. This section describes the process to collate public information across industries and countries/regions, and the design and administration of the online survey used to explore businesses' use of Facebook apps and technologies.

3.1 PUBLIC DATA

3.1.1 DATA COLLECTION

The first step in the economic analysis was to collate public economic data for all countries and regions. Consistent with the method for calculation of GVA and exports, this consisted of:

- Employment by industry
- GVA by industry
- Output by industry
- Exports in terms of total goods and total services

Efforts were made to collate each component of economic data from a single source. This was desirable both from a consistency perspective and also for expediency reasons. However, this was not always possible for GVA and Output. Where needed, data is converted to US dollars using foreign exchange rates from the World Bank, IMF, and RBA. The following provides a simplified list of the public data sources used in the analysis, with a detailed breakdown in Appendix A.1:

- United Nations and OECD – GVA and output
- Australian Bureau of Statistics (ABS) – for statistics on Australia
- Department of Statistics, Malaysia (DOSM) – for statistics on Malaysia
- National Statistics, Republic of China – for statistics on Taiwan
- International Labour Organization (ILO) – employment
- World Bank – exports, purchasing power parity rates, and FX rates for most countries and regions
- Reserve Bank of Australia (RBA) – FX rates for some countries and regions not available with the World Bank
- International Monetary Fund (IMF) – GDP and FX rates for South Africa

In addition, output data for some countries and regions at an industry level was not available. Where this public data was not available, it was estimated using a linear regression. The linear regression used GVA values to estimate output for countries/regions and industries for which data was unavailable. There was a strong positive correlation between GVA and output, with an overall R-squared of 99.9 per cent across all industries and countries/regions.³⁰

Given the multiple data sources, a sense check was performed as to the consistency of the public data compiled for this analysis against a single data source (the World Input Output Database, WIOD).³¹ Whilst all 30 in-scope countries and regions are not included in the WIOD, for countries and regions with similar data available, this check demonstrated the datasets used in the economic analysis were broadly comparable.

³⁰ On an industry level regression, R-squared values ranged between 87.3 per cent and 99.97 per cent.

³¹ This was performed for the latest available data. Source: WIOD (2016), Socio economic accounts release 2016, published February 2018. Available at: <http://www.wiod.org/database/seas16> Accessed October 2019.

Table 3: Industry Classifications

Industry Code in the survey (19)	ISIC REV4 (21)	ILO Industries ³² (14)	REV4 Industry Name	ILO Industry Name
1	A	A	Agriculture, forestry, and fishing	Agriculture; forestry and fishing
2	B	B	Mining and quarrying	Mining and quarrying
3	C	C	Manufacturing	Manufacturing
4	D	D,E	Electricity, gas, steam, and air conditioning supply	Utilities
5	E	D,E	Water supply; sewerage, waste management, and remediation activities	Utilities
6	F	F	Construction	Construction
7	G	G	Wholesale and retail trade; repair of motor vehicles and motorcycles	Wholesale and retail trade; repair of motor vehicles and motorcycles
8	H	H,J	Transportation and storage	Transport; storage and communication
9	I	I	Accommodation and food service activities	Accommodation and food service activities
10	J	H,J	Information and communication	Transport; storage and communication
11	K	K	Financial and insurance activities	Financial and insurance activities
12	L	L,M,N	Real estate activities	Real estate; business and administrative activities
13	M	L,M,N	Professional, scientific, and technical activities	Real estate; business and administrative activities
14	N	L,M,N	Administrative and support service activities	Real estate; business and administrative activities
15	O	O	Public administration and defence	Public administration and defence; compulsory social security
16	P	P	Education	Education
17	Q	Q	Human health and social work activities	Human health and social work activities
18	R	R,S,T,U	Arts, entertainment, and recreation	Other services
19	S	R,S,T,U	Other service activities	Other services
20 ³³	T	R,S,T,U	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	Other services
21	U	R,S,T,U	Activities of extraterritorial organisations and bodies	Other services

³² International Labour Organization (2019), ILOSTAT - ILO modelled estimates: Employment by sector, Available at: <https://ilostat.ilo.org/data/bulk/> Accessed September 2019.

³³ Industries 20 and 21 (T and U) were not included as an option in the survey as this was likely to add confusion for the responders, therefore there were 19 available selections (industries A to S) within the survey.

3.1.2 INDUSTRY CLASSIFICATION

The economic analysis could only be completed based on the industry classification of publicly available data (i.e. the analytical tool could only operate based on the industry distribution of public data with the fewest industry classifications because it is possible to combine industries to match groupings, but it is not possible to dissect industries that are already grouped). The fewest industry classifications used in public data was for employment figures, which followed the ILO industry classifications as shown in Table 3 (column: ILO industries). The industries used in public data (ISIC REV4) were mapped to the ILO industry classification, detailed in Table 3.

3.1.3 CALIBRATING A BASE YEAR

Not all public data was available for the most recent calendar year (2018). Therefore, to enable consistent analysis any data prior to 2018 was extrapolated to 2018 levels. For example, of the 30 countries' and regions' GVA data, 17 were from 2018, 9 were from 2017, 3 from 2016, and 1 from 2015. The extrapolation method used the GDP growth rates from the IMF to extrapolate GVA and output from the most recent year of available data to 2018.³⁴ For countries and regions where we have more recent GVA data than output data (for example, 2018 GVA and 2017 output), the GVA ratio was calculated for 2017 and output for 2018 estimated as: GVA of 2018 divided by GVA ratio of 2017.

21

3.2 SURVEY

This section describes the design and administration of a survey that spanned businesses across all industries in the 30 countries and regions included within the study. The research team designed the survey, which was distributed and administered by Qualtrics as the selected platform provider. In accordance with data privacy laws in the 30 surveyed countries and regions, no personally identifiable information (PII) was collected during the research. The opt-in survey was distributed via email, with respondents offered a financial incentive for participating.³⁵

3.2.1 DEFINE SURVEY BOUNDARIES

Countries and regions surveyed

The survey was distributed to 30 countries and regions selected by the Facebook company, listed in Table 4.

³⁴ International Monetary Fund (2019), World Economic Outlook: Gross Domestic Product, current prices. Available at: <https://www.imf.org/en/Publications/WEO>. Accessed September 2019.

³⁵ The type of reward varied based on the panelist profile but included: cash, airline miles, gift cards, redeemable points, charitable donations, sweepstakes entrance, and vouchers. Value of incentives ranged from approximately \$5-10 (USD).

Table 4: List of Surveyed Countries and Regions

Argentina	Colombia	Italy	Pakistan	Taiwan, China
Australia	Egypt	Japan	Peru	Thailand
Bangladesh	France	Malaysia	Philippines	Turkey
Brazil	Germany	Mexico	Poland	United Kingdom
Canada	India	Netherlands	South Africa	United States
Chile	Indonesia	Nigeria	Spain	Vietnam

Industries covered

The survey used a self-reported measure to identify the industry to which the respondent's business (or employer) is most aligned. The options provided were consistent with the United Nations (UN) International Standard Industrial Classification (ISIC) Rev 4, which categorises industries into 21 classifications.³⁶ Two industries (T & U) that are marginal components of GVA, were removed from the available list to reduce complexity and cognitive load.³⁷ Therefore, there were 19 industry classification response options.³⁸ This includes the three public sectors that are not used within the economic analysis, as the survey aimed to gather samples from all industries. To help respondents select the most appropriate industry, a definitional description on each industry was provided.

Sampling strategy

The survey aimed to uncover how businesses used Facebook apps and technologies across all industries and the in-sample countries/regions. In each country or region, employees and business owners were used as proxy informants for the business they worked for and were invited to participate in an opt-in online survey.³⁹

Characteristics of the business population for all 30 countries and regions and industries are unknown. To the knowledge of the research team, there are no consistent public sources for the count of businesses for each industry, or geodemographic data for all 30 in-scope countries and regions. This is especially true for the developing countries and regions in our sample, where online panels were the only way to access employees at the scale required. Therefore, the selection of a representative sample is extremely hard to achieve. The results from the survey

³⁶ United Nations (2008). International Standard Industrial Classification of All Economic Activities (ISIC), Rev 4. Page 58.

³⁷ Industry T is 'Activities of households as employers; undifferentiated goods and services producing activities of households for own use'. Industry U is 'Activities of extraterritorial organisations and bodies'. Together these industries account for less than one per cent of the total GVA for the in-scope countries and regions. It is expected that any survey respondents from these industries are likely to have reported they are within 'other service activities' which is within the same industry as T and U when grouped for analysis.

³⁸ While the survey provides a list of 19 industries for respondents to select from, due to employment data being provided for 14 industry classifications, the analytical tool operates on the basis of 14 industries (being the lowest common denominator). The mapping of the ISIC Rev 4 industries to ILO classification is provided in Section 3.1.1.

³⁹ Using employees as proxy informants is inferior to using auditable information regarding the business, but this is often not publicly available. Efforts were made to ensure survey responses were reasonable through screening out nonsensical respondents (i.e. outlier removal) and the inclusion of a survey question that screened out respondents with no input to decision making for their business (see Section 3.2.2 for further detail on screening questions).

and corresponding economic analysis should not be viewed as being perfectly representative of the economies covered (see Section 3.2.4 for more information on selection bias in the survey sample).

Quota sampling was used in an attempt to minimise selection bias; maximum quotas were set for each industry based on the distribution of GVA across industries within that specific country or region.⁴⁰ We chose to take a tailored sample in proportion to the population by distribution of GVA as GVA was the central research variable of the study.

Sample sizes

To ensure appropriate coverage across industries in each country or region, and subject to feasibility constraints, the survey was open until 500 qualified respondents in each country or region had responded. This provided a total sample of 15,342 responses, distributed across each country or region based on the distribution of GVA by industry.

The selection of a sample size of 500 per country or region was made by trading off:

- **The desire to select a sample that is representative of the population⁴¹** – the primary objective of the study was to estimate economic measures across 30 countries and regions, as such representativeness was sought at the country/region level. The survey provider, Qualtrics, estimated the desired sample size at 390 per country/region.⁴² As the true number of businesses in each country or region and the rate of Facebook app and technology use by business were unknown, the target sample size was increased to 500 responses per country or region to reduce the potential for error.
- **Practical and financial considerations** – as the economic measures are calculated at the industry level, consideration was given to having a sample size that is representative at a country/region and industry level, noting that true representation was impossible to achieve without access to consistent geodemographic and socioeconomic indicators. This was ruled out as a representative sample across 30 countries and regions and 11 industries (within the market economy) and was beyond the scope and resources of this research.

⁴⁰ GVA data by industry predominantly taken from UN Data and OECD. See Appendix A.1. for more information on public data sources.

⁴¹ The law of large numbers states that the observed sample average will approach the population average as sample size increases.

⁴² With the total count of businesses by industry being unknown for all countries and regions, an exact calculation for the required sample size was not possible. However, estimations were made based on an approximation of one million businesses per country/region, to derive a minimum sample size per country/region while assuming an incidence outcome variable (if binary) of 0.5 and a 95 per cent confidence level. This provides a sample size (n) of approximately 390 per country/region. We note that power calculations, like the one provided above, assume normal distribution of data. As financial metrics, such as total and international revenue, are positively skewed, the calculated sample size should be viewed as a conservative estimate. The survey targeted all businesses, and therefore collated responses for both on-platform and off-platform organisations.

Survey weighting

Survey weighting can be a useful tool to ensure that the sample chosen is as closely reflective of the population as possible when proportional sampling of the population is not possible. However, in order to use this technique appropriately, a rigorous understanding of the population is required.

As previously noted, there is no consistently available information on the number of businesses by industry and country/region, and therefore this view of the population is unavailable. Furthermore, consistent geodemographic data for each country and region is not available.

Therefore, a conservative approach, and one that does not introduce new bias, is to avoid survey weighting without a rigorous view of the size and demographics of the population. As such, these survey results should be considered as limited, and inference drawn on a larger population from these results should be done with due caution.

3.2.2 SURVEY DESIGN

The survey was designed to capture the relevant information regarding businesses across all industries in each of the sample countries and regions. No personal information was collected through the survey to comply with data privacy laws in each of the surveyed countries and regions. Respondents were invited to complete the survey in return for a small incentive. All respondents were free to terminate the survey at any point.

Survey translations

Six of the surveyed countries and regions were English-speaking nations, and 24 were not English-speaking. To ensure the accuracy of the translations for those who did not speak English, a two-stage, native speaker verification process was used. Through this process, the questions were translated into the required language by a native speaker, and these were then independently translated back into English using a secondary native speaker to test for accuracy.

Screening questions

Screening questions were used to ensure the respondent was currently employed and had input into decision-making for the business they were responding on behalf of to help increase the accuracy of survey data.

Furthermore, minimum sales limits were introduced after multiple nonsensical answers were received during the pre-test (e.g. \$16.15 (USD) and \$8.07 (USD) total sales value over the last 12 months). Based on analysis of the pre-test data and public data, a minimum annual total sales threshold of \$3,700 (USD) was set⁴³ and converted into local currencies using average exchange rates for 2018, as well as

⁴³ A value of \$3,700 (USD) was chosen based on a suitable value that prevented nonsensical responses from the pre-test data. As a reference point the research team found international precedent within the ABS' analysis of regional businesses where they do not include any businesses with total annual revenues below \$7,000 (USD). It was decided that a threshold of \$7,000 (USD) could prevent legitimate

responses, and therefore \$3,700 (USD) was a more sensible threshold based on the analysis of the pre-test data. As stated above, to take into account the real purchasing power across the in-sample countries and regions this was converted into reasonable country/region-specific thresholds using the PPP rates from the World Bank (2018).

the 2018 purchasing power parity values obtained from the World Bank.⁴⁴ This value was chosen based on analysis of the pre-test data to remove nonsensical responses without impacting the integrity of the analysis and removing legitimate micro-businesses. Additional outlier removal processes were used on the survey data for further cleansing of the results (covered in Section 3.2.5).

Facebook apps and technologies

The survey was designed to capture the overall impact of Facebook apps and technologies, rather than to attribute any perceived benefits to individual applications. Therefore, other than a question upfront which asked the respondent which Facebook apps and technologies they used, all further questions asked the respondent on behalf of all Facebook apps and technologies (whether the business used one or all four apps). A definition was provided on every page that asked for information regarding Facebook apps and technologies and was also included in the introductory pages of the survey.

Reverse-framed questions

Reverse-framed questions (i.e. set in terms akin to ‘willingness to accept’) were used to test for systematic response bias between positively and negatively framed questions that operationalise the percentage of sales supported by the Facebook company. On average, the negatively framed question was answered 5 percentage points lower than the positively framed question.

In line with the empirical literature on willingness to accept formulations of valuation questions, we have relied on the negatively framed question as the approach eliciting the most conservative valuation from the respondent.⁴⁵ Whilst this question cannot be portrayed as a basis to establish a fully causal link, we consider this as an approach to attribution that minimises positive response bias.

25

3.2.3 SURVEY TESTING (PRE-TEST)

An undeclared pre-test of the survey collected an average sample of 30 responses per country or region. The objective was to test the effectiveness of the survey questions and to refine outlier rules. Specifically, the sample was tested for the following:

- Data quality
- Descriptive statistics and distribution of data
- Time to complete
- Additional sense check for counterfactual questions
- Cultural understanding and accuracy of translations

⁴⁴ World Bank (2019). Purchasing Power Parity. PA.NUS.PPP.

⁴⁵ Kamoen, N., Holleman, B., & van den Bergh, H (2013). Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*, 7(3), 181-189. Respondents are more likely to disagree with a negatively-framed question.

Minor adjustments were made to a number of the survey questions on the basis of the analysis of the pre-test responses. The pre-test responses were not included in the final survey responses.

3.2.4 ACCOUNTING FOR BIAS

Bias arises in surveys where the sample systematically over- or underestimates a population parameter. Even after comprehensive design, bias can arise. The challenge is how best to mitigate the risks of encountering bias and errors, and therefore achieve the most reliable sample given the constraints of the study.

The following sections address some of the potential biases that were accounted for in the design, conduct, and analysis of the survey.

Selection bias

The survey was administered via online panels. These panels are unlikely to represent the true distribution of businesses by key geodemographic variables (e.g. size, industry, country/region), introducing systematic error into sample estimates. As the distribution of relevant geodemographic variables in the population is not known in all countries and regions surveyed, it is not possible to correct for this bias in a valid and reliable way, e.g. by post-weighting to match known population characteristics.⁴⁶

Efforts were made to ensure the survey sample was representative of the target population through quota sampling, using a sampling plan that was closely aligned to the distribution of GVA by industry for each country or region, as this is a known population characteristic.⁴⁷ However, as quota sampling is a non-probabilistic sampling technique, potential remains for selection bias, with key sources of bias considered below.

- First, online panels are known to suffer from selection bias, in that all respondents have access to the internet. As a result, employees who do not have access to the internet were not observed in our sample. As the use of Facebook apps and technologies require internet connection (mobile, wireless, fixed line, or otherwise), the use of an online survey is biased towards businesses and employees that are more likely to be on-platform. This is likely to contribute to an overestimation of the economic impact of the Facebook company (see Section on the following page: Correcting for selection bias).⁴⁸

⁴⁶ We note that the results of similar surveys have applied post-weighting to responses according to known geodemographic characteristics of the online population in a given country or region. This correction is not applied in the current methodology (1) as its validity is dependent on the assumption that the business and online populations in any given country or region are homogenous; and (2) as this information is not available for all 30 in-scope countries and regions.

⁴⁷ We set sample quotas that matched the distribution of GVA for the 14 industries in each country or region. The density of responses from an industry was not allowed to exceed the quota by more than 3 percentage points.

⁴⁸ This is particularly true in countries and regions with limited internet penetration, where the difference between online and offline businesses is larger.

- Second, due to the approach of using employees as proxy informants of their employer's business activities, it is likely that large businesses are overrepresented in the survey sample, compared to large firms' share of the overall count of businesses.⁴⁹ However, linear regression suggests that business size (operationalised using number of employees) is inversely related to the probability of a business being on-platform, and the proportion of sales supported by the Facebook company. As a result, qualitative differences on these factors are not likely to contribute to an overestimation of the value Facebook apps and technologies contributes to the global economy.⁵⁰

Correcting for selection bias

A common statistical technique used to correct for selection bias is the Heckman two-step correction (the Heckman Correction).⁵¹ Conceptually, this is a type of correction tool that may be applied by explicitly modelling the individual sampling probability of each observation (the so-called selection equation) together with the conditional expectation of the dependent variable (the so-called outcome equation). Doing so in a way that adds to the consistency of estimates requires prior information on the nature of sampling bias – information which in this context is limited.

A key root of sampling bias is partial internet penetration, logically driving a direct overrepresentation of online businesses (as the offline businesses and employees are not sampled, by design). To inform the identification of the most appropriate selection equation, we reviewed the density of observed geodemographic characteristics. This analysis indicated that in countries and regions with poor internet penetration, respondents were more likely to be from on-platform businesses, as well as: founders or leaders of their businesses, from businesses with a founder who is under 25 years old, and from businesses with a founder who is university-educated. Given this positive association, we identified a potential risk of overestimating the economic impact of the Facebook company in countries and regions with lower internet penetration which, in turn, are mostly in developing economies.⁵²

In an effort to correct for this bias, we used a selection equation that corrected for the tendency of the business founder being in sample, when looking at the effect of being on-platform, industry, country/region, business type, and business classification on total sales.⁵³ All the model specifications had poor predictive power at the explanatory stage – likely due to insufficient information since available internet penetration data is at the country/region level and does not vary by industry. Thus, the Heckman

⁴⁹ In countries and regions where public data on the share of employment and firms by business size is available, our in-sample share of responses by business size is more closely aligned to the share of businesses by employment, than to the equivalent shares measured with respect to the overall count of firms.

⁵⁰ Businesses with many employees are likely to have a greater value of sales; since sample estimates are scaled using GVA for the industry and country/region selected, the risk of bias due to over-representation of large businesses is mitigated.

⁵¹ Heckman JJ (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, 5(4), 475–492.

Toomet, O. & Henningsen (2008). Sample Selection Models in R: Package sampleSelection. *J. Stat Softw* 27(7):1–23. <http://www.jstatsoft.org/v27/i07/>.

⁵² This is a post-hoc approach to defining the selection equation, which can increase the likelihood of detecting a false positive (through model over-fitting). While this is not the preferred approach, the analysis was undertaken, in an attempt to correct for the observed deficiencies in our sample. We note that the presence of the observed selection bias in developing countries and regions leads to concerns about bias in itself.

⁵³ Selection equation: Founder ~ business age + business headcount + industry + country/region. Outcome equation: Total sales ~ On-platform + industry + country/region + business type (goods vs services vs both) + business classification (B2B vs B2B vs both).

Correction could not yield reliable results. For this reason, we have set aside this approach and, as stated above, we applied a more direct approach to correct for varying levels of internet penetration across the surveyed countries and regions.

Other biases

Retrospective reporting, or recalling past information through surveys, can give rise to measurement error. People can develop memory biases and gaps in information over time, which can distort the data captured, and reduce the precision of calculated estimates of the Facebook company's economic impact.⁵⁴ A number of steps were undertaken to minimise measurement error, including:

- The survey design minimised cognitive load for both question comprehension and memory recollection (e.g. survey respondents were provided with an option to report business data for a period of time they were comfortable with (i.e. 1-12 months), which was converted to an annual figure for the purposes of the economic analysis).
- The measurement instrument was pilot-tested to identify and correct potential response biases.
- Multiple measures were included for key constructs (e.g. including both positive and negatively framed questions to quantify the proportion of sales supported by Facebook apps and technologies).

Furthermore, survey data (as stated above) was set up to provide financial metrics over the last 12 months, October 2018 – October 2019. In contrast, the public data used referred (or was extrapolated) to the 2018 calendar year (1 January 2018 – 31 December 2018). As a result, there is a small discrepancy in the time period covered by public and survey data. We do not expect this discrepancy to be significant, and, in lieu of reliable public data for year-to-date 2019, no further action was taken.

3.2.5 SCREENING OF OUTLIERS

Survey data scrubbing

Data scrubbing refers to the procedure of modifying or removing incorrect, inaccurately formatted, or repeated data in a dataset, with the objective to ensure that the data is accurate and consistent for further analysis. This process happens before survey cleansing (covered in the next section), which removes outliers from the dataset.

⁵⁴ The confidence intervals around the base measures (such as total and international sales) were extremely wide due to large variations seen by country/region and industry. We note that, when used in calculations, due to error propagation, the magnitude of measurement error increases, decreasing the precision of calculation results. In addition, the methodology used to estimate the

contribution of the Facebook company to the global economy required sample estimates to be scaled up to represent the size of the economy of a given industry in a given country/region. This further decreases the precision of economic estimates provided.

Qualtrics analysed survey responses to help identify bad responders. This identified:

- Responders attempting to game the system through multiple responses.
- Bots (machines) automatically completing the survey.
- Responders providing contradictory responses.

Post scrubbing, there were a total of 15,342 responses.

Survey cleansing

Due to the right-tailed distribution of the data, classical outlier removal methods (e.g. removal of values more than three standard deviations above or below the mean) were not applicable as they removed significant proportions of the data. These methods are more applicable for data that is normally distributed.⁵⁵ A further method to remove outliers based on the median absolute deviation was also discounted as it had the unintended consequence of removing responses that were deemed reasonable. Therefore, a rule-based approach to removing outliers was used, based on the feasibility of responses.

The approach to removing outliers based on feasible possibility requires assumptions regarding the point at which a value becomes infeasible.⁵⁶ The assumptions taken in our approach are outlined below:

- Respondents should provide their responses in a currency commonly referenced in the country or region they have selected (i.e. local currency or US dollars in a country outside Europe; or local currency, US Dollars, GB Pound, or the Euro where the response is from a country within Europe).
- Headcount should not exceed that of the largest public company in the country or region selected.
- Total sales growth should not exceed 100,000 per cent.
- Total sales should not exceed either:
 - The total output (public data) for the selected industry in the selected country/region, or
 - The total sales of the largest public company in the selected country/region (public data).
- Headcount growth should not exceed 100,000 per cent.

⁵⁵ As total sales have a floor of \$3,700 (USD) and no ceiling, and international sales figures have a floor of \$0 (USD) and no ceiling, this result was expected. The observed distribution of total sales and total international sales are positively skewed, with the sample median being less than the mean.

⁵⁶ Public data for Output and Exports for each industry in each country/region, as referenced in Section 3.1. Largest company in each country/region: Forbes (2019). The World's Largest Public Companies, <https://www.forbes.com/global2000/list/#header:country>. Accessed September 2019.

Maximum sales growth: the figure for maximum sales growth (100,000 per cent) draws on a number of public sources, all of which cite maximum sales growth (over a 3 year period) as being less than 100,000 per cent, for example:

- 51,364 per cent – Financial Times (2019). FT1000: Europe's fastest growing companies, www.ft.com/content/238174d2-3139-11e9-8744-e7016697f225. Accessed September 2019.
- 36,680 per cent – Inc (2019). Inc. 5000: The Most Successful Companies in America, <https://inc.com/inc5000/2019/top-private-companies-2019-inc5000.html>. Accessed September 2019.

- International sales should not exceed the total exports for the type of business (i.e. total services exports if they are in the services sector and likewise for total goods exports) in the country or region selected (public data).
- The total value of sales supported by a single employee should not exceed \$10,000,000 (USD).
- A respondent should not indicate that they are not a business leader or founder, and yet also claim to have sole decision-making authority.

Responses that violated any of these assumptions were removed from our sample, resulting in the removal of 732 responses. The total number of complete survey responses received was 15,342, of which 12,750 were in the industries associated with the market economy. After the removal of infeasible responses, the sample size was 14,505, of which 12,018 were in the market economy.

In instances where a respondent reported international sales greater than total sales, but had passed the prior outlier checks, a value for international sales was imputed using the smaller of either the respondents' total sales, or the average international sales for the selected industry in the selected country/region.



AP

PENDIX

A.1 PUBLIC DATA SOURCES

Table A: Legend for Data Sources

Employment	GVA	Output	Exports
ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
	UNDATA, National Accounts Official Country Data, data not consistent with the most recent year available for other countries and regions	UNDATA, National Accounts Official Country Data, data not consistent with the most recent year available for other countries and regions	World Bank, National Accounts data and balance of payments statistics, data not consistent with the most recent year available for other countries and regions
	OECD, National Accounts, data for 2018 year	OECD, National Accounts, data for 2018 year	Taiwan, China specific: National Statistics, Annual IO tables, data for 2017 year
	OECD, National Accounts, data not consistent with the most recent year available for other countries and regions	OECD, National Accounts, data not consistent with the most recent year available for other countries and regions	
	Australia specific: ABS, Australian System of National Accounts, data for 2018 year. Also, OECD National Accounts data for 2017 year for financial services industry.	Australia specific: ABS, Australian Industry, data for 2018 year. Also, estimated for financial services industry using regression from GVA.	
	Taiwan, China specific: estimated using regression	Taiwan, China specific: National Statistics, National Accounts, data for 2017 year	
		Malaysia specific: Department of Statistics Malaysia, data for 2015 year for Real Estate, Education, and Health. Estimated for all other industries. Data not available: estimated using regression from GVA data	

Table B: Summary of Data Collected, Source, and the Latest Year of Data Available

Country/Region	Employment	GVA	Output	Exports
Argentina	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year. Also, estimated for some industries using regression from GVA data.	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Australia	ILO, ILOSTAT, data for 2018 year	ABS, Australian System of National Accounts, data for 2018 year. Also, OECD, National Accounts, data for 2017 year for financial services industry.	ABS, Australian Industry, data for 2018 year. Also, estimated for financial services industry using regression from GVA.	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Bangladesh	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2017 year. Estimated 2018 output using GVA ratio for 2017. Also, estimated for some industries using regression from GVA data.	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Brazil	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2016 year	UNDATA, National Accounts Official Country Data, data for 2016 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Canada	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2015 year	OECD, National Accounts, data for 2015 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Chile	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2017 year	OECD, National Accounts, data for 2017 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Colombia	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2017 year. Estimated 2018 output using GVA ratio for 2017.	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Egypt	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2017 year	UNDATA, National Accounts Official Country Data, data for 2015 year. Estimated 2017 output using GVA ratio for 2015. Also, estimated for some industries using regression from GVA data.	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
France	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2018 year	OECD, National Accounts, data for 2018 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Germany	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2018 year	OECD, National Accounts, data for 2018 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
India	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2016 year	UNDATA, National Accounts Official Country Data, data for 2016 year. Estimated output for the utilities industry (D,E) using the GVA ratio of electricity (D) for 2016.	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Indonesia	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	Estimated using regression from GVA data	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)

Country/Region	Employment	GVA	Output	Exports
Italy	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2018 year	OECD, National Accounts, data for 2018 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Japan	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2017 year	OECD, National Accounts, data for 2017 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods) and 2017 year (services)
Malaysia	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2017 year	Department of Statistics Malaysia, Annual Economic Statistics 2018 and Economic Census 2016, data for 2015 year for Real Estate, Education, and Health industries. Estimated 2017 output for those industries were estimated using GVA ratio of 2015. Estimated using regression for all other industries.	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Mexico	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2017 year	OECD, National Accounts, data for 2017 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Netherlands	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2017 year	OECD, National Accounts, data for 2017 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Nigeria	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2017 for agriculture and 2016 for all other industries. Estimated 2018 output using GVA ratio for 2017 for agriculture and 2016 for all other industries.	World Bank, National Accounts data and balance of payments statistics, data for 2017 year (goods and services)
Pakistan	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Peru	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Philippines	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	Estimated using regression from GVA data	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Poland	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2017 year	OECD, National Accounts, data for 2017 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
South Africa	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year and some industries for 2017. The education and health industries are reported as a combined industry, which has been included in education.	UNDATA, National Accounts Official Country Data, data for 2017 year. Estimated 2018 output using GVA ratio for 2017. Also, estimated for some industries using regression from GVA data.	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)

Country/ Region	Employment	GVA	Output	Exports
Spain	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2018 year	OECD, National Accounts, data for 2018 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Taiwan, China	ILO, ILOSTAT, data for 2018 year	Estimated using regression	National Statistics, National Accounts, data for 2017 year	National Statistics, Annual IO tables, data for 2017 year
Thailand	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	Estimated using regression from GVA data	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
Turkey	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2018 year	Estimated using regression from GVA data	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
United Kingdom	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2016 year	OECD, National Accounts, data for 2016 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods and services)
United States	ILO, ILOSTAT, data for 2018 year	OECD, National Accounts, data for 2017 year	OECD, National Accounts, data for 2017 year	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods) and 2017 year (services)
Vietnam	ILO, ILOSTAT, data for 2018 year	UNDATA, National Accounts Official Country Data, data for 2018 year	Estimated using regression from GVA data	World Bank, National Accounts data and balance of payments statistics, data for 2018 year (goods) and 2016 year (services)

A.2 NUMBER OF SURVEY RESPONSES BY COUNTRY/REGION AND INDUSTRY

The following table shows the number of survey responses by country/region and industry (for the market economy sectors only) after the screening of outliers is applied.

Table C: Count of Survey Responses by Country/Region and Industry (Post Outlier Screening)

Industry Country/Region	A	B	C	D,E	F	G	H,J	I	K	L,M,N	R,S,T,U	Total
Argentina	10	8	60	11	42	44	51	19	34	74	35	388
Australia	12	14	39	11	53	40	45	19	28	99	30	390
Bangladesh	40	2	94	18	25	18	66	10	32	50	68	423
Brazil	13	5	38	13	37	44	53	16	44	95	29	387
Canada	13	12	56	8	37	17	59	20	49	97	23	391
Chile	21	20	27	13	51	48	52	23	29	98	26	408
Colombia	17	8	51	26	72	28	66	18	32	97	26	441
Egypt	16	4	89	24	44	29	60	11	32	69	19	397
France	11	3	59	17	30	33	44	12	30	70	29	338
Germany	5	10	102	13	27	34	59	17	31	50	33	381
India	26	2	93	17	32	23	66	17	40	72	24	412
Indonesia	27	21	116	9	41	39	65	11	35	40	25	429
Italy	14	4	68	13	40	16	49	6	31	49	108	398
Japan	3	0	111	12	40	43	64	9	27	64	34	407
Malaysia	17	6	130	11	41	42	58	27	46	36	31	445
Mexico	7	9	78	8	55	45	55	25	31	83	26	422
Netherlands	10	3	71	7	26	46	62	6	41	54	30	356
Nigeria	45	8	31	19	33	24	72	20	34	69	35	390
Pakistan	38	6	69	18	23	27	73	26	24	71	25	400
Peru	21	22	47	4	47	34	66	31	25	62	31	390
Philippines	27	1	64	13	46	35	43	24	52	75	25	405
Poland	10	5	113	12	51	35	66	10	35	51	27	415
South Africa	25	19	68	25	36	48	63	21	63	69	20	457
Spain	15	1	61	11	40	24	48	13	29	95	34	371
Taiwan, China	6	3	127	15	10	42	47	23	22	67	29	391
Thailand	24	3	139	20	30	42	39	40	18	33	24	412
Turkey	10	8	93	15	62	17	53	22	28	53	26	387
United Kingdom	3	5	60	13	37	22	68	14	47	90	35	394
United States	5	2	66	7	28	28	59	17	54	96	28	390
Vietnam	35	3	98	21	43	44	34	18	29	48	30	403
Total – All countries and regions												12,018

Table D: Industry Codes

Code	Industry
A	Agriculture; forestry and fishing
B	Mining and quarrying
C	Manufacturing
D,E	Utilities
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H,J	Transport; storage and communication
I	Accommodation and food service activities
K	Financial and insurance activities
L,M,N	Real estate; business and administrative activities
R,S,T,U	Other services

Sources:

- Australian Bureau of Statistics (2019), *5204.0 – Australian System of National Accounts, 2018–19*. Retrieved from: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/7d12b0f6763c78caca257061001cc588/bfef5c43f53f8d2bca257213001d23c9!OpenDocument>
- Australian Bureau of Statistics (2019), *8155.0 – Australian Industry, 2017–18, Business size by Industry Division*. Retrieved from: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/8155.02017-18?OpenDocument>
- Department of Statistics Malaysia (2018), *Annual Economic Statistics 2018, Education Main Findings*. Retrieved from: https://www.dosm.gov.my/v1/index.php?r=column/ctHEMEByCat&cat=328&bul_id=NIJ4L1JWKzIMZTImSERVdS8vMUpJUT09&menu_id=b0pIV1E3RW40VWRTUkZocEhyZ1pLUT09
- Department of Statistics Malaysia (2018), *Annual Economic Statistics 2018, Health and Social Work Main Findings*. Retrieved from: https://www.dosm.gov.my/v1/index.php?r=column/ctHEMEByCat&cat=329&bul_id=NmEvSmNCOTZSWnRGWC82ZDhPZIN6Zz09&menu_id=b0pIV1E3RW40VWRTUkZocEhyZ1pLUT09
- Department of Statistics Malaysia (2016), *Economic Census 2016, Real Estate Main Findings*. Retrieved from: https://www.dosm.gov.my/v1/index.php?r=column/ctHEMEByCat&cat=417&bul_id=RTNrS2xRZDUzWXR0bGtnWIRSZWJZQT09&menu_id=b0pIV1E3RW40VWRTUkZocEhyZ1pLUT09
- International Labour Organization (2019), *ILOSTAT – ILO modelled estimates: Employment by sector*. Retrieved from: <https://ilostat.ilo.org/data/bulk/>
- International Monetary Fund (2019), *Exchange Rate Report*. Retrieved from: <https://www.imf.org/external/np/fin/ert/GUI/Pages/CountryDataBase.aspx>
- International Monetary Fund (2019), *World Economic Outlook: Gross Domestic Product, current prices*. Retrieved from: <https://www.imf.org/en/Publications/WE0>
- Organisation for Economic Cooperation and Development (2019), *National Accounts – Dataset: 6A. Value added and its components by activity*. Retrieved from: https://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE6A
- National Statistics, Republic of China (Taiwan), (2018), *National accounts, Statistical tables: Gross Domestic Product by Kind of Activity and Implicit Price Deflators*. Retrieved from: <https://eng.stat.gov.tw/ct.asp?xItem=37408&CtNode=5347&mp=5>
- National Statistics, Republic of China (Taiwan), (2018), *Annual IO tables for Commodity-by-Industry: 2017 Transactions Table at Purchasers' Prices*. Retrieved from: <https://eng.stat.gov.tw/ct.asp?xItem=29540&ctNode=1650&mp=5>
- Reserve Bank of Australia (2019), *Historical Data, Exchange Rates*. Retrieved from: <https://www.rba.gov.au/statistics/historical-data.html>
- United Nations Data (2018), *National Accounts Official Country Data: Table 2.1 – Value added by industries at current prices (ISIC Rev. 3)*. Retrieved from: http://data.un.org/Data.aspx?d=SNA&f=group_code%3A201
- United Nations Data (2018), *National Accounts Official Country Data: Table 2.3 – Output, gross value added, and fixed assets by industries at current prices (ISIC Rev. 3)*. Retrieved from: http://data.un.org/Data.aspx?d=SNA&f=group_code%3A203
- United Nations Data (2018), *National Accounts Official Country Data: Table 2.4 – Value added by industries at current prices (ISIC Rev. 4)*. Retrieved from: http://data.un.org/Data.aspx?d=SNA&f=group_code%3A204

United Nations Data (2018), *National Accounts Official Country Data: Table 2.6 – Output, gross value added and fixed assets by industries at current prices (ISIC Rev. 4)*. Retrieved from: http://data.un.org/Data.aspx?d=SNA&f=group_code%3A206

World Bank (2019), *International Monetary Fund, Balance of Payments Statistics Yearbook and data files*. Retrieved from: <https://data.worldbank.org/indicator/bx.gsr.mrch.cd>

World Bank (2019), *International Monetary Fund, International Financial Statistics*. Retrieved from: <https://data.worldbank.org/indicator/PA.NUS.FCRF?view=chart>

World Bank (2019), *World Bank national accounts data and OECD National Accounts data files*. Retrieved from: <https://data.worldbank.org/indicator/NE.EXP.GNFS.CD>



Copenhagen
Economics

CE