NBER WORKING PAPER SERIES

ANATOMY OF TECHNOLOGY AND TASKS IN THE ESTABLISHMENT

Xavier Cirera
Diego A. Comin
Marcio Cruz

Working Paper 32281
http://www.nber.org/papers/w32281

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2024

We thank David Baqaee, Mary Hallward-Driemeier, John Haltiwanger, Maurice Kugler, Bill Maloney, Martha Martinez Licetti, Denis Medvedev, Dani Rodrik, Giacomo Ponzetto, Doug Staiger, Edouard Schaal, Chris Snyder, Eric Verhoogen, and seminar participants for insightful comments. Kyung Min Lee actively participated in earlier drafts of this paper and we are indebted to him. Harneet Singh provided outstanding research assistance. We thank Aman Mahajan, Magda Malec, Mariana Pereira, Santiago Reyes and Antonio Martins Neto for inputs to sections of the paper. We also thank the industry experts and partner institutions for their support in designing and implementing the survey. An extended list of acknowledgements is provided in the appendix. Financial support from the infoDev Multi-Donor Trust Fund, the Korea World Bank Group Partnership Facility (KWPF), and the Competitive Industries and Innovation Program (CIIP) is gratefully acknowledged. This manuscript supersedes Cirera et al.(2021) and Cirera et al. (2020). The views expressed in this paper are solely those of the authors and do not necessarily reflect those of the World Bank Group, its Board, or the National Bureau of Economic Research.

Anatomy of Technology and Tasks in the Establishment
Xavier Cirera, Diego A. Comin, and Marcio Cruz
NBER Working Paper No. 32281
March 2024
JEL No. O33

ABSTRACT

We construct a grid that covers the key business functions of an establishment and the main technologies used in each of them. We populate this grid with data from over 20,000 establishments in 15 countries. We use this dataset to document novel "facts" about how establishments use technology, the sourcing of business functions, the specialization of establishments from a task perspective, the measurement of technology, and the relationship between technology sophistication and productivity across establishments. We find that differences in technology sophistication account for 31% of cross-establishment dispersion in productivity and for more than half of the agricultural productivity gap.

Xavier Cirera
The World Bank
1818 H ST NW
Washington, DC 20433
United States
xcirera@worldbank.org

Diego A. Comin
Dartmouth College
Economics Department
6106 Rockefeller Hall, Room 327
Hanover, NH 03755
and CEPR
and also NBER
diego.comin@dartmouth.edu

Marcio Cruz
IFC, World Bank Group
2121 Pennsylvania Avenue, NW
Washington, DC 20433
marciocruz@ifc.org

1 Introduction

Technology is central to some of the most fundamental questions in economics. Yet, our understanding of these matters largely relies either on indirect or on very limited measures of technology. In this paper, we develop a new approach to directly and comprehensively measure the technology of establishments. We apply our methodology to assemble a dataset that covers over 20,000 establishments in 15 countries and document how establishments use technology and the relationship between technology and productivity.

A long tradition in economics and sociology going back to Ryan and Gross (1943) and Griliches (1957) has characterized technology in the establishment by the presence of a few (typically one) advanced technology.[1] This approach faces several limitations. First, the number of technologies covered is very small when compared to the number of technologies involved in production in the average establishment. Second, the focus on the presence of advanced technologies makes it impossible to understand how production takes place in establishments that do not utilize state-of-the-art technologies. This concern is particularly relevant in developing countries where advanced technologies are less widely diffused. Third, traditional measures omit how intensively a technology is used in the establishment. As a result, we are ignorant about whether the most widely used technologies in establishments are the best technologies they have available and about the relative importance for productivity of the presence of a technology vs. how intensively it is used.[2]

To overcome these limitations, we develop a two-dimensional grid structure (henceforth, the grid) that conceptualizes the use of technology in the establishment. On the vertical dimension, the grid covers the key groups of tasks (referred to as "business functions") that an establishment conducts. For each business function (BF), the grid lists the key technologies that the establishment can use to conduct the relevant tasks.[3] The grid covers 7 business functions that are common across all sectors, which we call general business functions (GBF), and 56 BFs that are specific to one of 12 large sectors[4] and that we call sector-specific business

[1]Since their classic work on hybrid corn in agriculture, many have applied this approach to other technologies and sectors. For example, Davies (1979) studies the diffusion of 26 different manufacturing technologies, each technology is typically relevant in only a single narrow sector, Trajtenberg (1990) measures the presence of CAT-scanners in hospitals, Brynjolfsson and Hitt (2000); Stiroh (2002); Bresnahan, Brynjolfsson and Hitt (2002); Akerman, Gaarder and Mogstad (2015) measure the presence of some ICTs such as computers or access to internet. More recently, Acemoglu et al. (2022) focused on five generic, frontier technologies: AI, robotics, dedicated equipment, specialized software and cloud computing.

[2]See Comin and Hobijn (2007), Comin and Mestieri (2018).

[3]We assembled the grid with the help of over 50 experts that are knowledgeable about the activities involved in production in each sector as well as the technologies required to conduct them. Their names are listed in the acknowledgements section of the Appendix.

[4]Agriculture (crops), livestock, food processing, apparel, leather goods, automotive, pharmaceutical, other manufacturing, wholesale and retail, financial services, land transport services, and health services.

functions (SSBF). In total, the grid spans 305 technologies.

The grid has three properties. First, it is comprehensive both in terms of the business functions and of the technologies considered in each business function. Second, it is relevant for any establishment and country, regardless of its level of development. Third, for a given business function, the technologies in the grid can be ranked according to their sophistication, from the simplest to the most complex which represent the world technology frontier.

We operationalize the grid in an establishment survey; the Firm Adoption of Technology (FAT) survey. FAT collects three types of information. First, it collects establishment-level information on sales, inputs, education of the workers and the managers including management practices, etc. Second, for each sector-specific business function of each establishment, it records whether the SSBFs is conducted in-house, outsourced to another firm or in-sourced to another establishment of the firm. Third, FAT records all the technologies from the grid used by the establishment in each business function conducted in-house and, among those, which is the most widely used technology in each business function.

We have administered the FAT survey to representative samples[5] of establishments in 15 countries which include South Korea, Poland, Croatia, Chile, the Brazilian state of Ceará, Georgia, Vietnam, the Indian states of Uttar Pradesh, Tamil Nadu, Gujarat and Maharashtra, Ghana, Bangladesh, Kenya, Cambodia, Senegal, Ethiopia, and Burkina Faso. The resulting dataset comprises over 20,000 establishments.

The information collected with FAT allows to construct measures of the number of different technologies used by an establishment in a business function (NUM), and the sophistication of the most widely used (MOST) and the most sophisticated technologies used in a business function (MAX). We use these measures and other information collected with FAT to document novel "facts" about how establishments use technology at the business function level, the sourcing of business functions, the specialization of establishments from a task perspective, the measurement of technology, and the relationship between technology sophistication and productivity across establishments.

Zooming at the business function level, we document that establishments simultaneously use multiple technologies of different sophistication in a given business function; that the most widely used technology typically is NOT the most sophisticated an establishment has in the business function; and that the difference between the identity of the MAX and MOST technologies is not transitory. We note that these findings contradict core predictions of quality ladders (Aghion and Howitt, 1992) and love-for-variety models (Romer, 1990) which are the canonical theoretical treatments of technology use in the firm. These frameworks have emphasized that the best available technology at the business function level (MAX) is

[5]FAT is representative at national, regional, sector, and establishment-size levels.

a sufficient statistic for the vector of technologies involved in a business function. Instead, the data highlights that MAX and MOST capture different technology upgrading processes and that MOST is key to characterize the technologies used in a business function.

The granular dissection of establishments all the way to the business function level, places FAT in a privileged position to study the scope of establishments from a task perspective. We document that, on average, establishments conduct an overwhelming majority of business functions in-house. This observation validates the approach of measuring the technological sophistication of an establishment by focusing only on the technologies used in functions conducted in-house. We also quantify the frequency of sourcing (outsourcing and in-sourcing) for all the sector-specific business functions in the grid extending prior work that typically has covered a few tasks in a few countries.[6] The literature on the limits of the firm going back to Coase (1937) has emphasized the role of asset specificity and contractual frictions for sourcing decisions (Grossman and Hart, 1986; Hart and Moore, 1986; Antràs, 2003; Nunn, 2007). FAT allows us to go beyond these classical determinants and study the association between sourcing and technology sophistication. We document that sourcing has an inverted U-shaped relationship with establishment-level technology sophistication. Additionally, sourcing is negatively associated with the gap between MAX and MOST in the establishment.

Aggregating the measures of technology sophistication across all the functions of an establishment yields establishment-level measures that inherit the comprehensiveness of the grid. We explore how much of the cross-establishment variation in technology sophistication is missed by traditional measures of technology that reflect the presence of specific technologies such as computers, electricity, internet access, ERPs or industrial robots, or by narrow measures of technology sophistication that reflect the technology sophistication in only one business function. We show that the former explain a small part of the cross-establishment variance in technology sophistication, while the explanatory power of the latter depends much on the identity of the business function, making it quite risky to infer establishment-level technology sophistication from detailed technology information from just one function.

A quintessential question in the literature is the relation between technology and productivity.[7] Estimating productivity regressions, we show that technology sophistication is strongly, and robustly associated with productivity. Cross-establishment differences in tech-

[6]For example, fabrication of specialized intermediate goods in the US (Fort, 2017), accounting in manufacturing SMEs in Belgium (Everaert, Sarens and Rommel, 2010), innovation and R&D in manufacturing in Japan (Ito, Tomiura and Wakasugi, 2007), and Belgium (Veugelers and Cassiman, 1999)

[7]These studies typically consider a limited number of technologies. For example, Hubbard (2003) focuses on on-board computers in trucks, Bartel, Ichniowski and Shaw (2007) on computer numerically controlled (CNC) machines and computer-aided design (CAD) software, Hjort and Poulsen (2019) on high-speed internet, Gupta, Ponticelli and Tesei (2020) on cellphones.

nology sophistication account for 31% of the differences in productivity. This finding shows that the difficulties encountered in the literature to find a strong association between technology and productivity (including the Solow paradox) reflect the failure to construct technology measures that cover both the considerable number of technologies involved in production as well as their actual use.

The productivity regressions yield three, additional, important insights. First, MOST is much more relevant than MAX for establishment productivity, underscoring even further the vital importance of bringing MOST to the center-stage of technology measurement and modelling. Second, there is large variation across sectors in the share of cross-establishment dispersion in productivity accounted for by technology sophistication. While in agricultural establishments it accounts for 50%, in services it accounts for 28%. As a result, differences in technology sophistication account for more than half of the agricultural productivity gap across establishments in high- vs. low-income countries (Caselli, 2005). Third, we explore whether technology is generically appropriate to use in both high- and low-income economies (Basu and Weil, 1998; Acemoglu and Zilibotti, 2001) by comparing the elasticity of productivity with respect to technology sophistication in high- vs. low-income countries. This elasticity is not smaller in the low-income subsample suggesting that the productivity gains from using more sophisticated technologies are not specific to advanced economies.

The rest of this paper is organized as follows. Section 2 introduces the FAT survey. Section 3 presents the technology measures and documents the use of technology at the business function level. Section 4 documents the frequency of sourcing of business functions to other establishments, the specialization of functions conducted in an establishment and describes the relationship between technology sophistication and both sourcing and specialization. Section 5 describes technology sophistication at the establishment level and compares the comprehensive measures we construct from FAT to traditional technology measures. Section 6 studies the relationship between productivity and technology sophistication across establishments. Section 7 concludes.

2 The Survey

The FAT survey (henceforth, “the survey") collects detailed information for nationally representative samples of establishments in agriculture, manufacturing, and services about the technologies that each establishment uses to perform key business functions necessary to operate in its respective sector. In the following sub-sections, we describe in detail the survey design and implementation.[8]

[8]See Appendix A for more details.

2.1 Structure

The survey is composed of five modules. Module A collects information on the general characteristics of the establishment.[9] Modules B and C cover the technologies used. Module D focuses on barriers to, and drivers of, technology adoption, while Module E gathers information about the establishment's balance sheet and employment.[10]

The survey differentiates between general business functions (Module B), which comprise tasks that all establishments conduct, regardless of the sector where they operate, and sector-specific business functions (Module C), which are potentially relevant only for establishments in a given sector. All establishments in our sample respond to Module B, but only those belonging to the sectors for which we have developed a sector-specific module respond to C. To attain a wide coverage that allows a meaningful study of sector-specific technologies, we develop sector-specific modules for 12 significant sectors in the economy.[11] These sectors have been selected based on their share in aggregate value-added, employment and number of establishments and cover all three industries (agriculture, manufacturing, and services).[12]

2.2 Technology grid

To design Modules B and C, we determined the business functions covered and the list of technologies that can be used to implement the key tasks in each function. We call the resulting structure "the grid".

To construct the grid, we followed three steps. First, we conducted desk research reviewing the specialized literature. Second, we held meetings with World Bank Group experts on each of the sectors covered. Third, we reached out to external consultants with significant experience (at least 15 years) in a given sector. For example, the external experts in agriculture and livestock were agricultural engineers and researchers from Embrapa-Brazil. For food processing, apparel, automotive, pharmaceuticals, transportation, finance, and retail, as well as for the GBFs, we relied on senior external consultants selected by a large management consulting organization. For health, our team relied on consultants and physicians with practical experience in both developing countries and advanced economies. In total,

[9] The survey is designed, implemented, and weighted at the establishment level. For multi-establishment firms, the survey targets the establishment randomly selected in the sample.

[10] The survey can be downloaded in the following address (https://dcomin.host.dartmouth.edu/files/FAT_Survey_complete.pdf).

[11] Agriculture, livestock, food processing, apparel, leather goods, automotive, pharmaceutical, other manufacturing, wholesale and retail, financial services, land transport services, and health services.

[12] The granular information that can be obtained with the FAT survey allows us to explore central questions on technology policy in developing countries. One example, itself a product of this paper, is the World Bank policy report "Bridging the Technological Divide" (Cirera, Comin and Cruz, 2022).

more than 50 experts participated in the construction of the technology grid. The resulting grid is composed of 7 general and 56 sector-specific business functions and contains a total of 305 technologies (See Section A.1.1 of the appendix for details on the procedures followed to define the grid). It is important to stress that all the technologies in the Grid are precisely defined so that their use can be objectively established by respondents and enumerators. Figure 1 presents the general business functions considered in the survey and the possible technologies that can be used to conduct each of them. Figure 2 presents the grid for agriculture-farming.[13]

In addition to identifying the key business functions and relevant technologies, experts also ranked the technologies associated with each business function based on their sophistication. The sophistication of a technology reflects the complexity of that technology and is often associated with its relative novelty (e.g., crypto payments are more complex than cash). More sophisticated technologies may be able to perform a wider variety of tasks or tasks of greater complexity, or may perform a given task with greater accuracy and speed, but do not necessarily imply greater productivity. Thus, the sophistication rankings are not based on ex-ante or ex-post information about the relative productivity associated with each technologies.

2.3 Information collected in FAT

The survey collects information in three broad areas: the business functions conducted by an establishment, the use of technologies in each business function, and information on the establishment's balance sheet, workers, and management.

2.3.1 Business functions

The business functions that comprise the vertical dimension of the grid cover the key tasks involved in production. Explorations conducted at the piloting stage of the survey as well as the responses to the questions on the use of technologies in GBFs demonstrate that these functions are conducted in-house and that respondents are aware about the technologies their establishments use in the GBFs.[14] We formally explore the relevance of each sector-specific business function in each establishment through a screener question that asks whether a sector-specific function is conducted in that establishment and, in case it is not, whether it

[13]The grids for the GBFs and the eleven SSBFs are available in Section A.1.1 of the appendix.

[14]Due to space constraints in the survey and the information revealed during the pre-pilot, we decided to not directly ask about whether establishments conduct each GBF in FAT. Proxying the fraction of GBFs that are not conducted in house by the share of GBFs for which the establishment responds that either "does not use" or "does not know if it uses" to all the technologies in the grid for the BF, we find that only 3.9% of GBFs are not conducted in-house.

is in-sourced, outsourced, or irrelevant. We use this information to construct measures at the business function and establishment levels that allow us to study the specialization of establishments from a task perspective and the limits of the establishment.

2.3.2 Technology questions

The survey has two types of questions about the technologies used to conduct each business function. First, it asks whether the establishment uses each of the technologies listed in the grid. After identifying the technologies that are used by the establishment in a business function, the survey asks which technology is the most widely used in that function. The answers to these questions permit us to differentiate between the range of technologies present in the business function vs. the intensity with which they are used.

FAT also asks whether the establishment uses "other technologies" in the business function in addition to those contained in the grid. Only in 3.65% of the business functions establishments declare that "other" technologies are used in the business function, and only in 0.8% of the business functions "other" is the most widely used technology. The low frequency of "other" demonstrates the comprehensiveness of the technologies in the grid.

In addition to collecting information on technology use based on the grid, the FAT survey also asks about the presence in the establishment of three general-purpose technologies: computers, internet, and electricity. The survey also includes other standard questions about balance sheet information, employment, education of the employees, and education and experience of the manager. The survey collects information on four management practices from MOPS.[15] The answers to these questions are used to construct a management z-score following the methodology in Bloom and Van Reenen (2007). Despite covering only four of the 16 variables collected in MOPS, the FAT z-score accounts for 90.5% of the cross-establishment variance of the original MOPS z-score for Mexican establishments collected by ENAPROCE.

2.4 The Data

Our analysis is based on primary data collected from establishments in 15 countries: South Korea, Poland, Croatia, Chile, Brazil (Ceará), Georgia, Vietnam, India (Uttar Pradesh, Tamil Nadu, Gujarat and Maharashtra), Ghana, Bangladesh, Kenya, Cambodia, Senegal, Ethiopia, and Burkina Faso. Several factors were considered in deciding where to implement the FAT survey. We targeted countries on different continents (Asia, Africa, South America,

[15]The four variables are presence of formal incentives, number of key performance indicators (KPIs), frequency of KPI review, and time frame of production targets.

and Europe), with different levels of income, for which there was access to a high-quality sampling frame. In these countries, we collected data from 21,055 randomly selected establishments from the sampling frames. Table 1 shows the distributions of the sample by country, sector, and size groups and Table C.13 provides descriptive statistics. The median establishment in our sample has 9 workers, with an average of 34 workers. 20% of workers have a college degree, 19% of firms were multi-establishments, 18% are part of a multinational firm, 17% are exporters, 18% are 5 years old or younger, and 76% have electricity, computers, and internet access.

2.4.1 Sampling

Our data is representative for a universe of about 2.1 million establishments.[16] The samples are nationally representative for establishments with 5 or more workers.[17] For each country, the sampling frame is based on the most comprehensive and up-to-date establishment-level census data available from the respective National Statistical Office (NSOs) or similar authority.[18] The survey is stratified on three dimensions - sector, firm size, and region - so that we can construct representative measures of technology for aggregates along these dimensions. Sampling weights are based on the inverse probability of selecting establishments within each stratum.

2.4.2 Measures to minimize bias and measurement error

The literature on survey design has identified three types of potential bias and measurement errors based on whether they originate from non-responses, the enumerator, or the respondent (Collins, 2003). In what follows, we briefly describe the steps taken in the design and implementation of the FAT survey to minimize these errors. Appendix A.4 provides a more detailed description of the measures implemented to minimize potential bias.

Non-response bias. To maximize response rates and minimize potential biases associated with non-response (Gary, 2007), we followed best practice procedures. First, we partnered with national statistical offices and industry associations to use the most comprehensive and updated sampling frame available. Second, we hired data collection companies or agencies which were supported by endorsement letters from local institutions and which

[16] Table A.2 provides information about the distribution of firms by country, sector, and size groups within the universe covered by the FAT survey.

[17] For the state of Ceará in Brazil, and the Indian states of Tamil Nadu, Uttar Pradesh, Gujarat and Maharashtra it is representative at the state-level.

[18] Section A of the Appendix provides more details on the sampling frame, survey implementation and data collection, and sampling weight.

had demonstrable experience in nationally representative firm-level surveys.[19] Third, we followed a standard protocol in which each firm was contacted several times to schedule an interview. Fourth, we mostly used mostly face-to-face or phone interviews, which usually have higher response rates than web-based interviews.[20]

Enumerator bias and error counts. The survey, training, and data collection processes were designed to minimize enumerator biases and data collection errors. First, we used closed-ended questions to make coding the answers a mechanical task, thereby eliminating the need for the enumerator to interpret the answers or exercise subjective judgement when coding them. Second, the same standardized training was implemented in each country in the local language, with enumerators, supervisors, and managers leading the data implementation. Third, we conducted a pre-test pilot of the questionnaire in each country using establishments not included in the sample. Fourth, to attain greater quality control during the data collection process, enumerators recorded the answers via *Computer-Assisted Personal Interviews* (CAPI) or *Computer-Assisted Telephone Interviewing* (CATI) software,[21] and we regularly monitored the data collection process using standard algorithms to analyze the consistency of the data.[22]

Respondent bias. We took several steps to minimize respondent bias. First, we ensured that the interview was arranged with the appropriate person or persons; main managers (and other managers, such as plant managers and HR managers, in larger firms). Second, we used a closed-ended design in the questionnaire such that the respondent was questioned about specific technologies one at a time and was not told beforehand all the technologies that were associated with each business function. This design reduced measurement error in respondent's answers. Third, we pre-tested the questionnaire in each country to ensure that our questions were clearly worded within the specific geographical and cultural contexts of each country, reducing the need for subjective judgement in responses (Bertrand and Mullainathan, 2001). Fourth, to avoid *social desirability bias*, which may cause respondents to overstate the use of more sophisticated technologies, the survey avoided the words "technology" and "sophistication", employing more neutral terms such as "methods" and "processes" instead.

[19]These procedures are in line with suggestions of good practice for implementation by (Bloom et al., 2016).

[20]The exceptions were Georgia and Croatia, were we used online surveys. In Georgia, we partnered with the National Statistical Office, which resulted in exceptionally high response rate. Note that face-to-face interviews were not possible during the pandemic and that the survey was implemented by telephone in some countries, as shown in Table A.3.

[21]For Georgia and Crotia, we used Computer Assisted Web Interviewing (CAWI) with online tools.

[22]Randomized survey experiments with household surveys have demonstrated that a large number of errors observed in *Pen-and-Paper Personal Interview* (PAPI) data can be avoided with CAPI (Caeyers, Chalmers and De Weerdt, 2012).

2.4.3 Ex-post checks and validation exercises

We conducted several ex-post checks to assess the quality of the collected data.

Non-response bias. The average (unit) response rate on the survey varies by country and ranges between 15% and 86%. For example, the response rate was 80% in Vietnam, 57% in Senegal, 39% in Ceará, Brazil, 24% in Korea, and 15% in Croatia. These response rates are high relative to typical response rates in firm-level surveys, which are around 5 to 10% and are consistent with response rates observed for WMS which are around 40% (Bloom et al., 2016).[23] To minimize potential non-response bias, we adjusted the sampling weights for unit non-response. The adjustment was calculated at the strata level, so that the weighted distribution of our respondent sample across strata (sector, size, region) exactly matches the distribution of establishments in the sampling frame.[24]

We conducted three tests to assess potential biases from unit non-response-rates.[25] In each of these exercises, presented in Section A.5 of the Appendix, we find no statistical difference in the number of employees, technological sophistication, wages, and share of workers by skill and education between firms in the group that proxies for the response sample and the group of firms that proxies for the non-response sample.

Response bias. To assess the relevance of response bias, we conducted a parallel pilot in Kenya where we re-interviewed 100 randomly selected firms with a short version of the questionnaire. For those firms, we randomly selected three business functions and asked about the presence of the relevant technologies.[26] We estimated a probit model to assess the likelihood of consistent answers between the original and the back-check interviews, controlling for firm-level fixed-effect. Reporting the use of a technology in the back-check interview is associated with 80.6% of likelihood of reporting the use of the same technology in the original interview. Conversely, reporting that a technology is not used in the back-check interview, is associated with a 70.7% likelihood of not being reported in the original

[23] Table A.4 in the Appendix A provides the response rate by country, defined as the ratio between firms that responded to the survey and the total number of eligible firms in the sample for which we attempted to conduct an interview. The response rates were higher when the survey was implemented by national statistical agencies.

[24] See Section A.3 of the appendix for more details on sampling weights.

[25] First, using the information from the sampling frame, we check if there are differences in the average number of workers per establishment between respondents and non-respondents within stratum. Second, using information on the number of contact attempts, we compare the establishment-level technology sophistication in GBFs, described in the next section, between establishments with above and below the average number of attempts. Third, in a similar vein, we compare establishments in the first list of contacts provided to interviewers, versus those provided subsequently. See Table A.5 to A.11 in Appendix A.

[26] The re-interviews produced 1,661 answers, 106 interviews times 3 business functions times an average of 5.2 technologies per function. Both the original and back-end interviews in the pilot are conducted by phone by different interviewers.

survey.[27] These estimates do not differ between establishments of different size.

Validation using external sources. We evaluate the quality and reliability of the data collected by comparing it to external sources in Korea (KED) and Brazil (RAIS).[28] We focus on variables related to establishment size, productivity and technology. Table A.11 shows that the weighted sample averages of the labor variables in the FAT data (number of workers, average wages, share of college workers, share of low- and high-skill workers) are not statistically different from the averages in the universe of firms from the RAIS dataset. FAT measures of log value-added per worker are strongly correlated with the log of average wages from RAIS (See Table A.10). In the Korean matched establishments we find very high cross-establishment correlations (above 0.93) in the log levels and growth rates of sales and employment, as well as in log labor productivity (0.73).[29] Additionally, the average adoption rate of ERP systems in Korean manufacturing establishments in FAT is similar to (Chung and Kim, 2021), who used a similar sampling frame (32% vs. 40% in Chung and Kim, 2021), and there is a strong cross-establishment association between the book value of machinery and equipment in KED and the establishment sophistication measures (MOST and MAX) from FAT, which will be explained in the next section.

Internal validation. We conduct an additional validation exercise of the technology measures, by studying whether establishments with larger sales, employment and sales per worker are more likely to use top-tier technologies.[30] Specifically, we estimate a linear probability model for each business function, where the dependent variable is binary and equal to 1 if the establishment uses one of the technologies classified as top-tier for the business function and 0 otherwise. The model includes a full set of country- and, for the GBFs, 2-digit sector fixed effects. The independent variables are either (log) sales, (log) employment or (log) sales per worker. We find that the coefficients for these variables are positive and significant in a large majority of business functions.[31]

These ex-post checks further reassure us about the soundness of the survey design, the

[27] The correlation between the binary responses in survey and pilot is 73% ranging from 65% in business administration to 77% in sales across business functions, and from 85% among the most basic technologies to around 61% in intermediate, and 77% at the most advanced technologies across functions.

[28] In Korea we merge FAT with the Korea Enterprise Data (KED), a leading supplier of business credit reports on Korean businesses. In Brazil, we merge the data with the *Relação Anual de Informações Sociais* (RAIS) which is an administrative database maintained by the Ministry of Labor providing information of salary for all formal workers in Brazil.

[29] The FAT survey asks about sales and number of employments for two periods. The most recent years for which the balance sheet is available are the year before the implementation of the survey and two years before that. For Korea, these reference years are 2019 and 2017.

[30] In Appendix A.1, we present in bold the technologies from each BF that are classified as top-tier.

[31] For sales we find a positive coefficient in 100% of BFs (85% significant at 5% level); for employment 98% are positive (93% significant); and for productivity 80% are positive (52% significant, and never negative and significant).

data collection process, and the accuracy of responses.

3 Technology in the business function

The horizontal dimension of the grid provides detailed information about which technologies are used in each business function of each establishment, and which of these technologies are used most intensively. With this information, we construct measures that help characterize the use of technology at the business function. These include the number of different technologies used in the business function, the most sophisticated technologies in the business function, the sophistication of the most widely used, and the presence of sophistication gaps among the different technologies used by an establishment in a business function. We then compare the patterns uncovered with the predictions of the canonical frameworks (quality ladders and love for variety models) used to model technology inside the firm.

3.1 Technology Measures

We denote by $ANUM_{f,j}$ the number of different technologies from the grid used in business function f in establishment j. When more than one technology is used in a business function, we explore whether the technologies used are contiguous in the sophistication ranking of the grid or, instead, there are sophistication gaps in the vector of technologies used. Formally, we define the sophistication gap of establishment j in business function f ($GAP_{f,j}$) as a binary variable that takes the value of 1 if the establishment uses technologies with sophistication rank τ and $\tau + k$ for $k \geq 2$ in function f but does not use the technology with sophistication rank $\tau + p$, for $1 \leq p < k$. $GAP_{f,j}$ is 0 when there are no gaps and at least two technologies are used in the function.[32]

We study the sophistication of the technologies used in a business function with two variables. $MAX_{f,j}$ measures the sophistication of the most sophisticated technology used in the given business function, while $MOST_{f,j}$ reflects the sophistication of the most widely used technology in the business function. The starting point to construct these measures is the experts' rankings of the technologies, from least to most advanced, $r_f \in 1, 2, ..., R_f$.[33]

[32] $GAP_{f,j}$ is not defined when less than two technologies are used in the function (i.e. $ANUM_{f,j} < 2$).

[33] Because several technologies may be assigned the same sophistication, the highest rank in a function R_f may be smaller than the number of possible technologies N_f. In a small number of business functions, the technologies covered are used in various subgroups of tasks. For example, in the body-pressing and welding functions of the automotive sector, the survey differentiates between technologies used for pressing skin panels, pressing structural components and welding the main body. In cases like this, we construct ranks of technologies for each subgroup of tasks within the business function, and then aggregate the resulting indices by taking simple averages across the tasks groups. See Appendix B.1.1 for more details.

We define the relative rank of a technology as $\hat{r}_f = \frac{r_f - 1}{R_f - 1}$. Note that $\hat{r}_f \in [0, 1]$. We follow the standard approach of constructing cardinal measures of the sophistication of a technology by applying an affine transformation to the relative rank, $\hat{r}_f$. In Section 6, we show that affine transformations are a reasonable cardinalization of ordinal technology measures because establishment (log) productivity is approximately linear in the cardinalized measures of technology sophistication.

Specifically, we define $MOST_{f,j}$ and $MAX_{f,j}$ as

$$MOST_{f,j} = 1 + 4 * \hat{r}^{MOST}_{f,j}. \tag{1}$$

$$MAX_{f,j} = 1 + 4 * \hat{r}^{MAX}_{f,j}, \tag{2}$$

where $\hat{r}^{MOST}_{f,j}$ and $\hat{r}^{MAX}_{f,j}$ are the relative sophistication rankings of the two technologies. By construction, $MOST_{f,j}, MAX_{f,j} \in [1, 5]$, and $MAX_{f,j} \geq MOST_{f,j}$. We also use a similar transformation to defined a scaled measure of the number of technologies used in a business function ($NUM_{f,j}$).[34]

Since the most sophisticated technologies in the grid define the current (world) technology frontier, $MAX_{f,j}$ and $MOST_{f,j}$ represent the closeness of a establishment to the technological frontier in a business function. $MAX_{f,j}$ and $MOST_{f,j}$ are of independent importance as they capture different aspects of the technology upgrading processes in the business function. $MAX_{f,j}$ increases when a firm implements a new technology that is more sophisticated than those currently in use in a given business function.[35] Therefore, increases in $MAX_{f,j}$ capture technology improvements as those in quality ladder (e.g., Aghion and Howitt, 1992) or horizontal variety (e.g., Romer, 1990) conceptualizations of technology in production.

Increases in $MOST_{f,j}$ occur because the establishment expands the use of a technology which becomes the new most widely used in the business function.[36] The new most widely used technology may be either a wholly new one in that business function or it may be a technology that was used marginally and whose use has been expanded. Therefore, $MOST_{f,j}$ is more closely connected to Mansfield (1963)'s notion of technology diffusion in the firm (in our case in the business function) than to innovation.

Relevant outcomes and observable characteristics are often reported at the establishment level. We construct establishment-level average technology measures as simple averages of $NUM_{f,j}$, $MAX_{f,j}$ and $MOST_{f,j}$ across the business functions of a establishment. Specifi-

[34]Formally, we define $NUM_{f,j}$ as $NUM_{f,j} = 1 + 4 * \frac{ANUM_{f,j} - 1}{N_f - 1}$, where N_f is the number of different technologies in the grid for business function f.

[35]This technology may not be new to the establishment, but it is new to the BF in the establishment.

[36]Obviously, for MOST to increase, the new most widely used technology must be more sophisticated than the previous most widely used technology.

cally, we define NUM_j, $MOST_j$ and MAX_j as:

$$S_j = \sum_{f=1}^{N_j} \frac{S_{f,j}}{N_j} \tag{3}$$

where $S = \{NUM, MOST, MAX\}$, and N_j is the number of business functions covered for establishment j.

3.2 Facts on technology use in the business function

We investigate the use of technology at the business function level from three different perspectives: the number of different technologies used in a business function; the presence of gaps in sophistication; and the comparison of the sophistication of the most sophisticated technology used in a function ($MAX_{f,j}$) with the sophistication of the most widely used technology ($MOST_{f,j}$). The patterns uncovered along each of these perspectives are then compared to those prescribed by canonical models of technology inside the firm.

Number of technologies

We document the number of technologies that an establishment uses in a given business function ($ANUM_{f,j}$). Column 1 of Table 4 reports the average of $ANUM_{f,j}$ across all the establishments that conduct a function, as well as the average across all SSBFs, all GBFs and all BFs. Table 3 reports the distribution of $ANUM_{f,j}$ across business functions/establishments. Then we focus on the function/establishment observations that use exactly one technology and study the distribution of the sophistication level of the technology used (Table 3). Fact 1 summarizes the key findings about the number of technologies used in a business function.

Fact 1.

A. On average, establishments use two different technologies per business function. This average is roughly the same for general and sector-specific business functions.

B. In 62.5% of business functions, establishments use more than one technology. In 52.5% of the BFs where only one technology is used, this is the least sophisticated in the grid.

Fact 1 shows that establishments use technologies very differently than what quality ladder models (e.g., Aghion and Howitt, 1992) predict. In contrast to quality ladders, most establishments use more than one technology per business function. In most cases where only one technology is used, it is not because the establishment has replaced a less sophisticated technology by a more sophisticated one, but because it has only adopted the least sophisticated technology in the grid.

Sophistication Gaps

We next explore the vector of technologies used in a business function. In particular, we explore the identity of the least sophisticated technology used in a function, and whether the technologies used are contiguous in the sophistication ranking. Table C.15 reports the fraction of establishments with sophistication gaps in each business function while Table 4 reports the average fraction across GBFs, SSBFs and all the business functions.

Fact 2.

A. In 70.4% of the business functions where an establishment uses more than one technology the establishment uses the least sophisticated technology in the grid.

B. Sophistication gaps occur in 25% of business function/establishment observations, and are more frequent in GBFs (27%) than in SSBFs (17%).[37]

Fact 2 shows that establishments do not tend to abandon dominated technologies (Part A) and that sophistication gaps are relatively rare. These two findings imply that the ranking of the most sophisticated technology used in a business function ($MAX_{f,j}$) is a good summary statistic for the vector of technologies used.

Technology sophistication

To better understand how establishments use technology, we explore the differences between the $MAX_{f,j}$ and $MOST_{f,j}$ measures of sophistication at the business function by conducting three exercises. First, we study the association between the two measures of sophistication as well as the association of each of these measures with the (scaled) number of technologies used in the business function ($NUM_{f,j}$). In particular, we estimate the following regressions:

$$MOST_{f,j} = \alpha_j + \alpha_f + \beta * MAX_{f,j} + u_{f,j} \tag{4}$$

$$MAX_{f,j} = \alpha_j + \alpha_f + \beta^{MAX}_{NUM} * NUM_{f,j} + u_{f,j} \tag{5}$$

[37]The GBFs where gaps are more frequent are payments (48%), business administration (34%) and sales (28%).

$$MOST_{f,j} = \alpha_j + \alpha_f + \beta_{NUM}^{MOST} * NUM_{f,j} + u_{f,j} \tag{6}$$

where α_j and α_f are establishment and business function fixed effects. (See Table 5 for the estimates.) Second, we compute a binary variable, $D_{f,j}$, that is 1 when $MAX_{f,j} > MOST_{f,j}$ and 0 otherwise.[38] Table 6 reports the fraction of business functions where $D_{f,j} = 1$ unconditionally, and conditional on $MAX_{f,j}$. Third, we study the sources of variation in $D_{f,j}$. To this end, we conduct a variance decomposition that determines the fraction of the variance accounted for by function and establishment effects (Table C.16). Then, we focus on the establishment component captured by the fraction of functions in the establishment where $MAX_{f,j} > MOST_{f,j}$ and study its association with establishment characteristics (Table 7). Fact 3 presents the key findings from these analyses.

Fact 3.

A. The average difference between $MAX_{f,j}$ and $MOST_{f,j}$ is 0.61. In 63% of the business functions where establishments use more than one technology, the most widely used technology is not the most sophisticated technology available.

B. $MAX_{f,j}$ and $MOST_{f,j}$ are positively correlated within establishments, however $MAX_{f,j}$ accounts for only 30% of the within establishment variation in $MOST_{f,j}$.

C. While an increase in $NUM_{f,j}$ by 1 is associated with an increase in $MAX_{f,j}$ by 0.85, it is associated with an increase in $MOST_{f,j}$ by only 0.25.

D. Observable establishment characteristics account for a small portion of the cross-establishment variation in the fraction of functions where $MAX_{f,j} > MOST_{f,j}$. The fraction of functions where $MAX_{f,j} > MOST_{f,j}$ is not related to the establishment's size and the associations with age, exporting, multi-national and multi-establishment status are quantitatively small.

Fact 3 shows that MAX and MOST measures capture distinct dimensions of the use of technology by establishments. Part A shows that, in most instances, the most widely used technology is not the most sophisticated that an establishment has available in a business function. Part B shows that even though the sophistication levels of MAX and MOST technologies are positively associated, they account for a small fraction of each other's variance.

Pact C sheds light on the technology upgrading process. Establishments increase the most sophisticated technology in a business function by bringing in technologies that are

[38] Because this dummy is trivially equal to 1 when only one technology is used in the business function, we restrict the analysis to business functions where an establishment uses at least two technologies.

new to the function. The estimate of β^{MAX}_{NUM} in (5) close to 1 confirms this premise. In contrast, the estimate of β^{MOST}_{NUM} in (6) is 0.25 suggesting that when establishments bring in a new technology they rarely make it the most widely used. These findings demonstrate that the dynamics of technology adoption are very different to the dynamics of expansion in the use of technologies in a business function (i.e. diffusion).

A natural question that arises is whether the gap between $MAX_{f,j}$ and $MOST_{f,j}$ is transitory, reflecting sluggishness in the diffusion of the use of new technologies in a business function of a establishment, [39] or whether the gap is permanent reflecting long-run costs of expanding the use of a technology or a strategic choice by the establishment.

Part D of Fact 3 sheds light on this question. The fact that the share of functions in an establishment with $MAX_{f,j}$ strictly greater than $MOST_{f,j}$ does not decrease much with the establishment's age suggests that the gap between these two measures is not a transitory phenomenon. To further explore this hypothesis we consider the subsample of function/establishment observations where the establishment has adopted one of the most sophisticated technologies in the function (that we denote as top-tier). Using information on a question in FAT that asks how long ago establishments adopted a top-tier technology, we split the sample between early adopters and recent adopters.[40] Figure 3 plots the histogram of $MOST_{f,j}$, for each of these subsamples. Conditional on adopting a top-tier technology, the distributions of $MOST_{f,j}$ for early and late adopters are quite similar. This confirms that the gap between $MAX_{f,j}$ and $MOST_{f,j}$ is largely permanent.

Establishment effects account for 29% of the variation in $D_{f,j}$,[41] but as stated in Fact 3C, only a small part of the variation across establishments in the gap between MAX and MOST is accounted for by observable establishment characteristics. What explains it then? One conjecture is that the establishment variation in this gap may reflect differences in technological strategies across establishments. Some establishments may chose to use as intensively as possible the most sophisticated technologies available in each function, while others may prioritize to bring in more advanced technologies over extending the use of existing technologies to a wider range of outputs/activities in the function. In the next section, we explore further the hypothesis that the average gap between MAX and MOST in the establishment has a strategic nature by connecting it to other strategic choices of the establishment such as the degree of specialization (from a task perspective) and the

[39] As in vintage capital models where establishments slowly replace obsolete technologies embodied in old capital as it depreciates (e.g., Benhabib and Rustichini, 1991).

[40] Appendix A lists the top-tier technologies in each business function. Early adopters are those that adopted a top-tier technology before the year of adoption of the median year of adoption of top-tier technologies in the specific business function. Recent adopters are those that have adopted after the median year of adoption.

[41] Business function effects account for 6%.

outsourcing and in-sourcing of business functions to other firms/establishments.

We conclude the discussion of the use of technology at the business function level by connecting our findings to the other canonical model of technology inside the establishment: the "love for variety model" (Romer, 1990). Consistent with "love for variety" models, we have documented that establishments simultaneously use multiple technologies in a business function and that $MAX_{f,j}$ is a good proxy for the range of technology sophistications used by an establishment in a business function. However, "love for variety" models completely ignore the intensity of use of technologies, which we show is important to understand the relationship between technology and productivity as well as the limits and specialization of the establishment, in the coming sections. Furthermore, the evidence that the most sophisticated technology is not the most widely used, is at odds with extensions of "love for variety" models where varieties differ in their productivity. In those extensions, the most widely used technologies in a business functions are the most productive ones which plausibly are also the most sophisticated. This implication is contrary to the distinct nature of the processes of adoption and diffusion of technologies at business function level documented in this section.

4 Business functions

The vertical dimension of the grid contains the business functions that are potentially relevant for an establishment. For SSBFs, FAT directly records whether a function is conducted in-house, in-sourced to another establishment of the firm, outsourced to another firm, or whether the business function is irrelevant for the establishment. [42] An analysis of this information is relevant for the study of technology for several reasons.

Sourcing allows establishments to indirectly access technologies available in other firms/ establishments. Therefore, ascertaining the frequency of sourcing is critical to assess the validity of establishment-level measures of technology that only cover in-house business functions. Most studies of sourcing are based on a small number of activities/functions in a few countries. FAT offers the possibility to document the outsourcing and in-sourcing of a comprehensive set of business functions, in 11 sectors and in 15 countries at very different stages of development. Many factors may influence an establishment decision of wither to source or not the activities that make up a business function. FAT allows us to study the potential role of in house technologies. Finally, FAT allows us to characterize the specialization of firms from the perspective of the diversity tasks/business functions conducted in-house.

[42]For GBFs, the pre-pilot together with the answers to the technology questions in FAT strongly suggest that GBFs are conducted in-house in an overwhelming majority of establishments. See Footnote 14

Specialization is typically studied from the perspective of the range of goods and services produced (e.g., Imbs and Wacziarg, 2003; Ekerdt and Wu, 2023).[43] Our data can shed light on the relationship between these two notions of specialization and their association with the technology sophistication of an establishment.

Measures of sourcing and specialization of establishments

To study the vertical dimension of the grid, we first define as relevant (sector-specific) business functions those that are conducted in-house, outsourced or in-sourced. Given this definition, we denote by IH_j the fraction of relevant (sector-specific) business functions an establishment conducts in-house, by O_j the fraction of relevant (sector-specific) functions that an establishment outsources to another firm, and by IN_j the fraction of relevant SSBFs that an establishment in-sources to another establishment in the same firm. Additionally, we note that we can measure the specialization of an establishment from a task perspective (SP_j) by the fraction of SSBFs that the establishment does not conduct in-house. That is, establishments that conduct a larger share of (sector-specific) functions in-house have a broader scope, and therefore are less specialized.[44]

Frequencies of outsourcing and in-sourcing and degree of task-based specialization

Table 8 reports the average of SP_j, IH_j, O_j and IN_j by sector and for the entire sample of establishments. The key finding is that 87% of the relevant sector-specific business functions are conducted in-house. Since an overwhelming majority of GBFs are also conducted in-house, this implies that the sophistication of technologies establishments use in-house (on average) is an accurate proxy for the sophistication of the technologies an establishment has access to both directly and indirectly (via sourcing of functions). On average, 11% of relevant SSBFs are outsourced and 2% are in-sourced. However, when conditioning on establishments in multi-establishment firms, the share of relevant SSBF in-sourced is 9%. Additionally, 24%

[43] There is a parallel literature that focuses on specialization of exports (Cadot, Carrère and Strauss-Kahn, 2011).

[44] See Figure B.16 for a detailed explanation of the relevant questions in the survey and the coding of each of the variables. In the appendix, we introduce binary measures that capture each of these dimensions at the business-function level. For example, $D^O_{f,j}$ is 1 if function f in establishment j is outsourced. These binary function-level measures allow us to conduct variance decompositions and study the contribution of establishment effects for the overall variation in outsourcing, in-sourcing or specialization of the function-level measures. Additionally, these measures can be used to construct establishment-level binary measures that reflect whether any of the SSBFs of a establishment are outsourced). In the Appendix we replicate most of the analyses using these extensive measures and find very similar results to those based on the intensive measures described above.

of SSBFs are not conducted in-house (either because they are sourced or because they are irrelevant for establishments).

The frequency with which business functions are sourced out of the establishment in FAT is in line with the literature. Fort (2017) documents that 27% of US establishments contract out the production of customized intermediate goods in other establishments (belonging to the same or different firms). Ito, Tomiura and Wakasugi (2007) show that 31% of firms in a sample of Japanese manufacturing companies outsource the production of components and other intermediates. They also study the outsourcing of non-production services which is less frequent than the contracting out of inputs. They report that 3% of Japanese manufacturing establishments outsource some R&D services, 6.7% some information services, 1.7% customer support, 5.9% professional services and 9.8% other tasks. We find that on average 9.1% (10%) of the relevant SSBFs are outsourced (sourced) in manufacturing and for Korea, which is the closest economy we have in our sample to Japan, on average 17.7% (20.3%) of the relevant SSBFs are outsourced (sourced).

To start exploring the sources of variation of SP_j, O_j and IN_j we compute the country and 2-digit sector components for each variable. The country effects account for between 1% and 8% of their cross-establishment variance, and the sector effects for between 5% and 8%.[45] With the caveat of the small number of countries in our sample, we study the relationship between the country effects and per-capita income. Across countries, specialization has a U-shaped association with per-capita income. This pattern is similar with the inverted U-shaped relation between product diversification and income documented by Imbs and Wacziarg (2003). The country effects of outsourcing and in-sourcing are positively associated with per capita income, although only the latter is statistically significant (see Table C.25).

Drivers of specialization, outsourcing and in-sourcing

To study the association between establishment characteristics and task specialization, and the outsourcing and in-sourcing of business functions, we estimate the following specification:

$$Q_j = \alpha_c + \alpha_s + \beta_1 * \overline{S}_j + \beta_2 * \overline{S}_j^2 + \kappa * (MAX_j - MOST_j) + \gamma * X_j + u_j \tag{7}$$

where $Q_j = \{SP_j, O_j, IN_j\}$, $\overline{S}_j$ is the average of MAX_j and $MOST_j$ in the establishment, and $MAX_j - MOST_j$ is the difference between the MAX and MOST measures of sophistication in the establishment. α_c and α_s are country and 2-digit sector dummies, and X_j is a vector of establishment dummies that capture age, size, and status as exporter, multi-

[45]See Table C.24 in the appendix. We also conduct a variance decomposition of the function-level counterparts and find that the establishment effects account for around 50% of their variance, while function effects account for around 10%.

national, and multi-establishment. Table 9 reports the estimates, and Fact 4 summarizes the key findings.

Fact 4.

A. Outsourcing (O_j), in-sourcing (IN_j) and task specialization (SP_j) have an inverted U-shaped relationship with establishment-level technology sophistication ($\overline{S}_j$), and are negatively associated with the gap between MAX and MOST in the establishment.

B. There is a strong positive association between task and product-based measures of establishment specialization.

The strong similarities in Fact 4 between the drivers of specialization, outsourcing and in-sourcing should not be surprising since an establishment that outsources or in-sources more SSBFs conducts fewer in-house and therefore it is more specialized from a task perspective.

Fact 4B shows that establishments that concentrate their sales on fewer products and services conduct fewer SSBFs in-house. However, the magnitude of the coefficient suggests that there are other important determinants of the fraction of functions conducted in-house.

The literature on the limits of the firm has emphasized the role of contractual frictions and asset-specificity as drivers of the limits of the firm/establishment (e.g., Grossman and Hart, 1986; Hart and Moore, 1986; Antràs, 2003; Nunn, 2007). Asset-specificity and contractual frictions are largely sector- and country-specific factors, respectively. However, the variance decompositions of O_j and IN_j (Table C.24) show that country and sector-specific factors capture a relatively small fraction of the observed variation in the intensity of outsourcing and in-sourcing by establishments.

Table 9 shows the relevance of age, exporting and multinational status for specialization and sourcing. Establishment's age is negatively associated with SP_j and O_j, but establishments between 6 and 15 years old in-source more SSBFs. There is negative association between exporting status and specialization and (to a lesser extent) with outsourcing. Being part of a multinational is positively associated with outsourcing but negatively with insourcing.

Fact 4A demonstrates the relevance of various dimensions of the technological sophistication of the establishment for its sourcing activities. The negative association between sourcing and the gap between MAX and MOST suggests that establishments may use the sourcing of business functions to have access to technologies more sophisticated than those available in-house.[46] The estimated relationship between $\overline{S}_j$ and sourcing in Table 9 implies that sourcing intensity is increasing with technology sophistication until $\overline{S}_j = 2.75$, which

[46] Admittedly, we do not observe what would be the sophistication of technologies an establishment would

corresponds to the 75^{th} percentile. This finding might reflect a complementarity between the sophistication of in-house technologies and those embodied in the goods and services from the sourced functions. Elucidating the specific channel behind the strong association between $\overline{S}_j$ and sourcing we have uncovered seems an important question for future research.

5 Technology sophistication across establishments

We move up from the business function to the establishment level to study technology sophistication across establishments. We are interested in two issues: The magnitude and sources of cross-establishment variation in technology sophistication, and the comparison of our measures of technology sophistication with traditional measures of technology in the establishment.

5.1 Sources of variation

We study the sources of variation in technology sophistication across establishments by (i) analyzing the distributions of MAX_j and $MOST_j$ (Table 10), (ii) computing the contributions of country and 2-digit sector effects to the cross-establishment variance in technology sophistication (Table C.17 and Table 11), and (iii) regressing technology sophistication on establishment characteristics such as age, size and status as exporter, multi-establishment and multinational (Table 12). Fact 5 summarizes the key findings.

Fact 5.

A. There is a large variation across establishments in technology sophistication. 23% (18%) of the variance of MAX_j ($MOST_j$) can be accounted for by country effects and 7% (4%) by 2-digit sector effects. The cross-country correlation between the country effects of MAX_j ($MOST_j$) and per capita income is 0.77 (0.93). The cross-establishment variance in technology sophistication is larger in agriculture, than services or manufacturing.

B. There is a positive association between technological sophistication and establishment size, being an exporter, part of a multi-establishment firm, or a part of a multinational firm. There is an inverted U-shape relationship between technology sophistication and the age of the establishment.

have access in-house for sourced functions but, Fact 3E has established that establishment effects account for a significant part of the variance in the gap between MAX and MOST for business functions conducted in-house. Therefore, the average gap in the establishment provides a good proxy for the hypothetical gap in the sourced function.

The large dispersion and high correlation with per capita GDP of both MAX_j and $MOST_j$ imply that technological sophistication is a relevant factor in a development accounting sense. MAX_j accounts for 38% and $MOST_j$ for 34% of the cross-country variance in per capita income.[47]

Fact 3A is very relevant for the literature on the agricultural productivity gap (e.g., Caselli, 2005; Lagakos and Waugh, 2013) that has documented the presence of larger cross-country differences in productivity in agriculture than in other sectors. The fact that cross-country differences in technological sophistication are also larger in agricultural establishments than in other sectors suggests that cross-establishment differences in technological sophistication may not only be relevant to account for cross-country differences in income per capita but also for the agricultural productivity gap. We explore this hypothesis in the next section.

5.2 On the measurement of technology at the establishment level

We next connect the technology sophistication measures from FAT to the literature on technology measurement at the establishment level. Specifically, we study how much of the cross-establishment variation in technology sophistication is missed by traditional measures of technology that are based on the presence of specific technologies, or by narrower measures of technology sophistication that cover only one business function. To this end, we regress $\overline{S}_j$ on various technology measures and study the R^2 of these regressions (see Table 13).[48] Fact 6 summarizes our key findings.

Fact 6.

A. Measures of the presence in the establishment of specific technologies, such as computers, electricity, and internet access, ERPs or industrial robots explain a small part $(7-30\%)$ of the cross-establishment variance in technology sophistication.

B. The fraction of cross-establishment variance in $\overline{S}_j$ accounted for the technological sophistication in a single business function depends much on the identity of the business function.

C. The technological sophistication of general business functions accounts for a larger share of the cross-establishment variance in average sophistication than the sophistication of sector-specific business functions. However, the relevance of technology in

[47] The contribution is computed as the covariance divided by the variance of per capita income. See Table 11.

[48] In the appendix we report estimates using MAX_j and $MOST_j$ as dependent variables.

SSBFs and GBFs for the variance of $\overline{S}_j$ varies across 1-digit sectors. The sophistication of SSBFs is most relevant in agriculture, while the sophistication of GBFs is most relevant in services.

Fact 6 shows the relevance of using comprehensive technological measures to accurately reflect the technological landscape of an establishment. Part A shows the limitations of the traditional approach to measuring technology which consists on reflecting the presence of a few specific technologies. A dummy that reflects the availability in the establishment of three general purpose technologies (computers, internet and electricity) just captures 18% of the cross-establishment variance in sophistication, while the presence of robots only accounts for 7% of the variance in sophistication across manufacturing establishments.

Part B highlights the risks of inferring the sophistication of an establishment from information in one business function. Technology sophistication in business administration forecasts 56% of the cross-establishment variance in technology sophistication. However, when using information on technology sophistication in payments we can forecast only 21% of the variance, and when using sophistication in fabrication, we can forecast 35% of the variance in the technology sophistication of manufacturing establishments. Part C of Fact 6 illustrates a related point: that the relevance of different types of functions (GBFs vs. SSBFs) differs across sectors. Given this variability in the relevance of business functions, we cannot be sure of how inaccurate our characterization of the technological landscape of an establishment is when relying on information from a limited number of business functions.

6 Technology and productivity

We conclude our deep dive into technology in the establishment by studying the relationship between technology and labor productivity. This relationship is central to important literatures in economics. The literature on the drivers of productivity across establishments and countries has studied why productivity differs so much between establishments in high- and low-income countries (e.g., Klenow and Rodríguez-Clare, 1997; Bartelsman, Haltiwanger and Scarpetta, 2013; Syverson, 2011). The literature on the agricultural productivity gap has studied why productivity differences are much larger among agricultural than non-agricultural establishments. The literature on appropriate technology has conjectured that establishments in low-income countries do not adopt advanced technologies because the scarcity of certain inputs (e.g., high-skilled workers, physical capital) prevent establishments to unlock their potential productivity gains and as a result, conclude that advanced technologies are inappropriate for low-income countries.

We contribute to these literatures by answering the following four questions: What fraction of cross-establishment differences in productivity can be accounted for by differences in technology? What dimensions of technology are more relevant for productivity? Is the contribution of technology sophistication to cross-establishment differences in productivity uniform across sectors? Is the association between technology sophistication and productivity different in high- and low-income economies?

To answer these questions, we estimate versions of the following establishment-level productivity regression:

$$Y_j = \alpha_s + \alpha_c + \beta_k * K_j + \beta_h * H_j + \gamma * S_j + \theta * X_j + u_j \tag{8}$$

where the dependent variable is the log of sales per worker, K_j is the log of the book value of capital per worker, H_j is the percentage of workers in the establishment with a college degree, S_j represents the measures of technology sophistication in the establishment, X_j is a vector of controls, α_s, and α_c are 2-digit sector and country dummies, and u_j is classical measurement error. Fact 7 reports the key findings in the relation between technology sophistication and productivity.

Fact 7.

A. Technology sophistication is strongly, and robustly associated with productivity. Cross-establishment differences in technology sophistication account for 28% of the differences in productivity, and for 20% of within-country differences in productivity. This contribution differs across sectors. In agriculture, technology sophistication accounts for 48% of differences in productivity (30% within countries), in manufacturing it accounts for 37% (27% within countries) and in services for 25% (17% within countries).[49]

B. Both MAX_j and $MOST_j$ are significantly associated with establishment productivity. The coefficient of $MOST_j$ is twice as large as the coefficient of MAX_j in the within-country specification, and roughly ten times larger in the specification that excludes country effects. Across business function, we find a stronger association of productivity with technology sophistication in general business functions than in sector-specific business functions.

[49] The contributions of technology sophistication to productivity differences are computed as follows. First, we compute the residual productivity for all firms by regressing productivity on the country and sector dummies, and the measures of physical and human capital, and then computing the residual. We do the same for $\overline{S}_j$. For each of these residuals, we calculate the gap between the 10^{th} and 90^{th} percentiles. We then multiply the coefficients of $\overline{S}_j$ in columns 1-2 and 6-11 of Table 14 times the 10-90 gaps in the residuals for each variable and divide the product by the residual of labor productivity. The resulting number is the percentage of the cross-establishment dispersion in productivity accounted for by technology sophistication.

C. Technology is generically appropriate to use in both rich and poor countries. The elasticity of productivity with respect to the technology sophistication of an establishment is not larger in high- than in low-income countries.

Robustness

As stated in Fact 7A, the cross-establishment relationship between technology sophistication and productivity is strong and robust. It is robust to including country, 2-digit sector and country-sector fixed effects (columns 1-3 of Table 14), to controlling for the quality of management practices (columns 4-5), and to allowing for sectoral variation in the elasticities of human and physical capital in the productivity regressions (columns 6-11).

Beyond its robustness, the coefficient of technology sophistication in the productivity regressions increases significantly after excluding the country fixed effects suggesting that differences in technology sophistication are even more relevant to explain cross-country than within-country differences in productivity.

Consistent with Bloom and Van Reenen (2007), we estimate a positive coefficient for management practices in the productivity regression (column 4). Its magnitude is relatively modest, and it accounts for 4% of the cross-establishment dispersion in productivity. We explore the possibility of a complementarity between technology sophistication and management[50] by introducing an interaction between $\overline{S}_j$ and a dummy that takes the value of 1 if the management score is above the median (column 5). We find that the coefficient is positive and significant, suggesting that a key role of managers is the proper implementation of more sophisticated technologies.[51]

Linearity

The productivity regressions shed light on whether the relationship between productivity and technology sophistication is approximately linear. This question is relevant for the validation of the cardinalization of the ordinal technology measures used to construct $MAX_{f,j}$ and $MOST_{f,j}$ in section 3. A natural cardinalization is to project the ordinal measures into establishment productivity so that there is a linear relationship between the (cardinal) technology sophistication measure and (log) productivity.

[50] Atkin et al. (2017) explore the role of organizational barriers and incentives in the adoption of superior cutting technology designs in the ball industry in Pakistan.

[51] In regressions reported in the online appendix, we have also controlled for other characteristics, such as exporter, multi-national and multi-establishment status. Our estimates are robust to these additional controls. We have also checked the robustness of the association to controlling for the measures of task-based specialization and sourcing presented in Section 4. Interestingly, we find that the coefficients of the measures of specialization and sourcing in the productivity regressions, likely reflecting the lower value added of the functions that more specialized establishments conduct in-house (See Table C.33).

The first two columns of Table 16 explore the linearity of the relation between $\overline{S}_j$ measure and productivity. Column 1 allows for a non-linear relationship by permitting the coefficient of $\overline{S}_j$ to differ across establishments ranked above or below the median sophistication level. Albeit the interaction between $\overline{S}_j$ and the "above median sophistication" dummy is negative and significant, its magnitude is small. Column 2 introduces greater flexibility by replacing $\overline{S}_j$ by dummies that classify establishments in four sophistication intervals.[52] The increments in productivity associated with a given increase in the average sophistication across intervals are roughly constant, confirming that the relation between productivity and $\overline{S}_j$ is well approximated by a straight line.[53] We therefore conclude that the linear cardinalization used to construct our sophistication measures is a good representation of the mapping from ordinal technology sophistication measures onto establishment productivity.

The Solow paradox

To explore the importance of the comprehensiveness of technology measures for the assessment of the impact of technology on productivity, we re-estimate the productivity regression (8) replacing technology sophistication ($\overline{S}_j$) by a less comprehensive measure such as the dummy for the presence of computers, internet and electricity in the establishment. The estimates are reported in Table C.31 and imply that the three general purpose technologies dummy accounts for 8% of the cross-establishment dispersion in productivity. This is three times less than what technology sophistication accounts for. Hence, the importance of comprehensive measures of technology not just to understand how establishments use technology but also the relationship between technology and productivity.

Sectoral heterogeneity

We estimate the productivity regression (8) separately for the three aggeregated one-digit sectors in order to (i) check the robustness of the relationship between productivity and technology sophistication to allowing for sectoral differences in country effects and in the elasticities of human and physical capital; and, (ii) explore the sectoral heterogeneity in the contribution of technology sophistication to cross-establishment productivity differences.

Columns 6 through 11 of Table 14 report the estimates. As stated in Fact 7A, the association between technology sophistication and productivity is positive and strong in all

[52]The intervals cover all the range of $\overline{S}_j$, and we require that a significant mass of establishments falls in each interval. Specifically, we use the following intervals: [1-1.5), [1.5-2.5), [2.5,3.5), [3.5,5].

[53]In particular, the increments are average $\overline{S}_j$ for each consecutive pair of intervals are 0.742, 0.881, 0.924. given the estimates in column 2 of Table 16, the increment in productivity per unit increments in sophistication are .35, .51, .465.

three sectors. There is significant sectoral heterogeneity in the coefficient of technology sophistication (γ), which is largest in agriculture and smallest in services. The higher coefficient for agriculture, together with the larger cross-establishment dispersion in technology sophistication (Table 10) explains the larger contribution of technology sophistication to the cross-establishment dispersion in productivity in agriculture than in the other sectors. This finding is very relevant for the agricultural productivity gap (Caselli, 2005).

The establishments in FAT sample display a very large agricultural productivity gap. The 10-90 cross-establishment gap in productivity is 1.91 log-points (or equivalently 6.75 times) larger in agriculture than in services.[54] This gap is considerable larger than the cross-country agricultural productivity gap which is around 2. Differences in technology sophistication account for a gap between the 10-90 cross-establishment productivity in agriculture vs. services of 1.05 log-points which is equivalent to a factor of 2.87. This implies that differences in technology sophistication account for more than half of the agricultural productivity gap.[55]

Technology dimensions

The multiple dimensions of technology that $\overline{S}_j$ can be decomposed to shed light on the relevance for productivity of different types of technology and technology upgrading processes (see Table 15). The simultaneous significant positive effect of MAX_j and $MOST_j$ in the productivity regressions (Fact 7B) suggests that bringing in new technologies and expanding the use of sophisticated technologies that were already adopted increase the establishment's productivity. However, the much greater coefficient, especially across countries, of $MOST_j$ suggests that increasing the use of technologies that were already adopted in the business functions (diffusion) is much more relevant for productivity than bringing in new technologies to the business function (innovation). This finding compels future research to adopt a shift in the paradigm used to connect technology and productivity in the establishment, as the canonical models (i.e., quality ladders and gains from variety models) focus exclusively on MAX_j.

Fact 7B also highlights the relevance of technology sophistication in GBFs and how the importance of technology in SSBFs varies across sectors. This finding highlights the potentially complex interactions between technology sophistication across the different business functions of an establishment. We plan to explore this question in future research.

[54] In agriculture it is 372, while in services it is 55.

[55] That is 1.05 vs 1.91.

Appropriate technology

The appropriate technology hypothesis has conjectured that rich countries tend to adopt more sophisticated technologies because their human and physical capital abundance enhances the productivity gains of new technologies (e.g., Basu and Weil, 1998; Acemoglu and Zilibotti, 2001). The comprehensiveness of FAT in terms of sectoral coverage, number and scope of technologies, as well as including establishments in both high and low income countries makes it ideal to explore whether technology is generically appropriate or not. To test this hypothesis, we split the sample between the three richer countries in the sample (South Korea, Poland and Croatia) that have higher average levels of human and physical capital, and the rest. Importantly, in both the high- and low-income subsamples, we have establishments at all levels of technology sophisticated. This allows us to explore whether the productivity gains associated with a higher sophistication of the technologies used in an establishment are lower in low-income countries, as the appropriate technology hypothesis has conjectured.

Columns 3-6 of Table 16 report the estimates from the productivity regressions allowing the coefficient on technology sophistication to vary between the high and low income countries.[56] In all specifications, we find that the coefficient of technology sophistication in the productivity regressions is not smaller for the sample of low-income countries than for the high income.

A valid concern at this point is that the estimates reflect the impact of productivity of an omitted variable that also affects the technology choices of establishments in developing countries. To start exploring this possibility, we make the plausible argument that omitted variables that affect differentially technology choices in low-income economies (such as access to finance, or scarcity of productive factors) are likely to affect the establishments' size. Therefore, we can learn about the role in our estimates of omitted variables vs. actual returns to technology sophistication by studying the association between technology sophistication and productivity after conditioning on establishment size. To this end, we split establishments by their size between those that are above and below the (unweighted) median number of employees which is 17.[57] We re-estimate the specification from column 3 of Table 16, in each of these two subsamples and report them in columns 7 and 8 of Table 16. There are several noteworthy observations. First, the elasticity of productivity

[56] Column 3 includes a dummy for the high income countries interacted with $\overline{S}_j$; columns 4 and 5 allow the coefficients on human and physical capital as well as the sectoral dummies to vary between the two groups of countries, and column 6 replaces $\overline{S}_j$ by four dummies based on the average sophistication of the establishment, and allows the coefficients to vary between the two groups of countries.

[57] This split provides an even distribution between large and small establishments in both low- and high-income countries.

with respect to capital per worker is similar across both samples, while the elasticity with respect to human capital is much greater for large establishments. Second, the estimate of the elasticity of productivity with respect to technology sophistication in the subsample of small establishments (column 8) is similar to the estimate for the overall sample (column 1), and it is significantly larger than for the subsample of large establishments (column 7). This finding suggests that the association between technology sophistication and productivity is unlikely to be driven by omitted variables that impact establishment size. Third, in both small and large establishments we find that the association between technology sophistication and productivity is not larger in high-income countries. Hence, we conclude that technology is generically appropriate to be used in all countries regardless of their development level. A corollary of this conclusion is that cross-establishment variation in technology sophistication must result from variation in the marginal costs of implementing and using more sophisticated technologies not in their marginal benefits.

7 Conclusions

This paper presents a new approach to comprehensively characterize the technologies used in an establishment based on a tool - the grid - that describes the key business functions involved in production and the possible technologies that can be used to conduct the main tasks in each function. We have implemented this methodology and assembled a dataset that covers over 20,000 establishments that constitute representative samples in 15 countries at all stages of development. An exploration of the FAT dataset has uncovered new facts about significant issues such as the use of technology at the business function level, the scope of business functions conducted in establishments, the measurement of technology at the establishment level and the relation between technology sophistication and productivity across establishments.

The facts documented in this paper underscore the value of new approaches to comprehensively measuring technology inside establishments. This value resides in both challenging existing frameworks about technology use and productivity and in opening the door for new questions that invite for the development of new theoretical paradigms. One of the key insights uncovered in this paper is the relevance of the MOST margin which not only has been ignored by the existing frameworks but they predict that should coincide with MAX. Therefore, we need richer frameworks to model the use of technology at the business function that that help us rationalize how establishments use of technology in the business function.

A second issue that needs new models is the aggregation of technology sophistication across business functions. In this paper, we have taken a reasonable shortcut for a descriptive exercise by constructing establishment-level measures of technology sophistication as the simple average of the function-level sophistication measures. However, developing a theory of aggregation that allows to construct establishment-level measures of technology sophistication directly from function-level information would be a major development.

Our exploration of the vertical dimension of the grid has uncovered strong associations between various dimensions of the technological sophistication of an establishment and its propensity to source business functions. These findings may rationalize the development of frameworks where the limits of the establishment do not just respond to contractual frictions but to technological strategies.

The fourth area that requires attention is the development of models that explain why the cross-establishment dispersion in technology sophistication is larger in agriculture, and the stronger association between productivity and technology sophistication for agricultural establishments.

References

Acemoglu, Daron, and Fabrizio Zilibotti. 2001. "Productivity Differences." *The Quarterly Journal of Economics*, 116(2): 563–606.

Acemoglu, Daron, Gary W Anderson, David N Beede, Cathy Buffington, Eric E Childress, Emin Dinlersoz, Lucia S Foster, Nathan Goldschlag, John C Haltiwanger, Zachary Kroff, et al. 2022. "Automation and the workforce: A firm-level view from the 2019 Annual Business Survey." *NBER Working Paper No. 30659.*

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323–351.

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad. 2015. "The Skill Complementarity of Broadband Internet." *The Quarterly Journal of Economics*, 130(4): 1781–182.

Antràs, Pol. 2003. "Firms, Contracts, and Trade Structure." *The Quarterly Journal of Economics*, 118(4): 1375–1418.

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit K. Khandelwal, and Eric Verhoogen. 2017. "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan*." *The Quarterly Journal of Economics*, 132(3): 1101–1164.

Autor, David H, and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review*, 103(5): 1553–97.

Bartel, Ann, Casey Ichniowski, and Kathryn Shaw. 2007. "How Does Information Technology Affect Productivity? Plant-level Comparisons of Product Innovation, Process Improvement, and Worker Skills." *The Quarterly Journal of Economics*, 122(4): 1721–1758.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. "Cross-Country Differences in Productivity: The Role of Allocation and Selection." *American Economic Review*, 103(1): 305–334.

Basu, Susanto, and David N. Weil. 1998. "Appropriate Technology and Growth." *The Quarterly Journal of Economics*, 113(4): 1025–1054.

Behaghel, Luc, Bruno Crépon, Marc Gurgand, and Thomas Le Barbanchon. 2015. "Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models." *The Review of Economics and Statistics*, 97(5): 1070–1080.

Benhabib, Jess, and Aldo Rustichini. 1991. "Vintage capital, investment, and growth." *Journal of Economic Theory*, 55: 323–339.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *American Economic Review*, 91(2): 67–72.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices Across Firms and Countries." *The Quarterly Journal of Economics*, 122(4): 1351–1408.

Bloom, Nicholas, Renata Lemos, Raffella Sadun, Daniela Scur, and John Van Reenen. 2016. "International Data on Measuring Management Practices." *American Economic Review*, 106(5): 152–56.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt. 2002. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." *The Quarterly Journal of Economics*, 117(1): 339–376.

Brynjolfsson, Erik, and Lorin M. Hitt. 2000. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." *Journal of Economic Perspectives*, 14(4): 23–48.

Cadot, Olivier, Céline Carrère, and Vanessa Strauss-Kahn. 2011. "Export Diversification: What's behind the Hump?" *Review of Economics and Statistics*, 93: 590–605.

Caeyers, Bet, Neil Chalmers, and Joachim De Weerdt. 2012. "Improving Consumption Measurement and Other Survey Data through CAPI: Evidence from A Randomized Experiment." *Journal of Development Economics*, 98(1): 19–33.

Caselli, Francesco. 2005. "Accounting for Cross-country Income Differences." *Handbook of Economic Growth*, 1: 679–741.

Chung, Sunghoon, and Minho Kim. 2021. "How smart is a 'smart factory'?: Drivers and effects of factory smartization."

Cirera, Xavier, Diego A Comin, Marcio Cruz, and Kyung Min Lee. 2020. "Technology within and across Firms." *Policy Research Working Paper, n. 9476. The World Bank.*

Cirera, Xavier, Diego A Comin, Marcio Cruz, and Kyung Min Lee. 2021. "Anatomy of Technology in the Firm." *NBER Working Paper Series, No. w28080. National Bureau of Economic Research.*

Cirera, Xavier, Diego Comin, and Marcio Cruz. 2022. *Bridging the technological divide: Technology adoption by firms in developing countries.* World Bank Publications.

Coase, R. H. 1937. "The Nature of the Firm." *Economica*, 4(16): 386–405.

Collins, Debbie. 2003. "Pretesting Survey Instruments: An Overview of Cognitive Methods." *Quality of Life Research*, 12(3): 229–238.

Comin, Diego, and Bart Hobijn. 2007. "Implementing Technology." *NBER Working Paper*, 12886.

Comin, Diego, and Martí Mestieri. 2018. "If Technology Has Arrived Everywhere, Why Has Income Diverged?" *American Economic Journal: Macroeconomics*, 10(3): 137–78.

Davies, Stephen. 1979. *The Diffusion of Process Innovations.* Cambridge University Press.

Ekerdt, Lorenz K.F., and Kai-Jie Wu. 2023. "The Rise of Specialized Firms." *Working Paper.*

Everaert, Patricia, Gerrit Sarens, and Jan Rommel. 2010. "Using Transaction Cost Economics to explain outsourcing of accounting." *Small Business Economics*, 35(1): 93–112.

Fort, Teresa C. 2017. "Technology and Production Fragmentation: Domestic versus Foreign Sourcing." *The Review of Economic Studies*, 84(2): 650–687.

Gary, Pike R. 2007. "Adjusting for Nonresponse in Surveys." In *Higher Education: Handbook of Theory and Research.* 411–449. Springer.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, 25(4): 501–522.

Grossman, Sanford J, and Oliver D Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94(4): 691–719.

Gupta, Apoorv, Jacopo Ponticelli, and Andrea Tesei. 2020. "Information, Technology Adoption and Productivity: The Role of Mobile Phones in Agriculture." National Bureau of Economic Research.

Hart, Oliver D, and John Moore. 1986. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98(6): 1119–1158.

Hjort, Jonas, and Jonas Poulsen. 2019. "The Arrival of Fast Internet and Employment in Africa." *American Economic Review*, 109(3): 1032–79.

Hubbard, Thomas N. 2003. "Information, Decisions, and Productivity: On-board Computers and Capacity Utilization in Trucking." *American Economic Review*, 93(4): 1328–1353.

Imbs, Jean, and Romain Wacziarg. 2003. "Stages of Diversification." *The American Economic Review*, 93(1): 63–86.

Ito, Banri, Eiichi Tomiura, and Ryuhei Wakasugi. 2007. "Dissecting Offshore Outsourcing and R&D: A Survey of Japanese Manufacturing Firms." *Research Institute of Economy, Trade and Industry (RIETI) - Discussion papers*, 07060.

Klenow, Pete, and Andrés Rodríguez-Clare. 1997. "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" *NBER Macroeconomics Annual*, 12.

Lagakos, David, and Michael E Waugh. 2013. "Selection, Agriculture, and Cross-country Productivity Differences." *American Economic Review*, 103(2): 948–80.

Mansfield, Edwin. 1963. "Intrafirm Rates of Diffusion of an Innovation." *The Review of Economics and Statistics*, 348–359.

Nunn, Nathan. 2007. "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade." *The Quarterly Journal of Economics*, 122(2): 569–600.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of political Economy*, 98(5, Part 2): S71–S102.

Ryan, B., and N. Gross. 1943. "The Diffusion of Hybrid Seed Corn in Two Iowa Communities." *Rural Sociology*, 8(1): 15–24.

Särndal, Carl-Erik, Bengt Swensson, and Jan Wretman. 1992. *Model Assisted Survey Sampling.* Springer Science & Business Media.

Stiroh, Kevin J. 2002. "Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say?" *American Economic Review*, 92(5): 1559–1576.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 2(49): 326–365.

Trajtenberg, Manuel. 1990. *Economic Analysis of Product Innovation: The Case of CT Scanners.* Vol. 160, Harvard University Press.

Ulyssea, Gabriel. 2018. "Firms, Informality, and Development: Theory and Evidence from Brazil." *American Economic Review*, 108(8): 2015–47.

Veugelers, Reinhilde, and Bruno Cassiman. 1999. "Make and buy in innovation strategies: evidence from Belgian manufacturing firms." *Research Policy*, 28(1): 63–80.

Williamson, Oliver E. 1979. "Transaction-Cost Economics: The Governance of Contractual Relations." *The Journal of Law and Economics*, 22(2): 233–261.

Williamson, Oliver E. 1988. "Technology and transaction cost economics: A reply." *Journal of Economic Behavior Organization*, 10(3): 355–363.

World Bank. 2009. "Enterprise Survey and Indicator Surveys: Sampling Methodology." *The World Bank, Manuscript.*

Figures

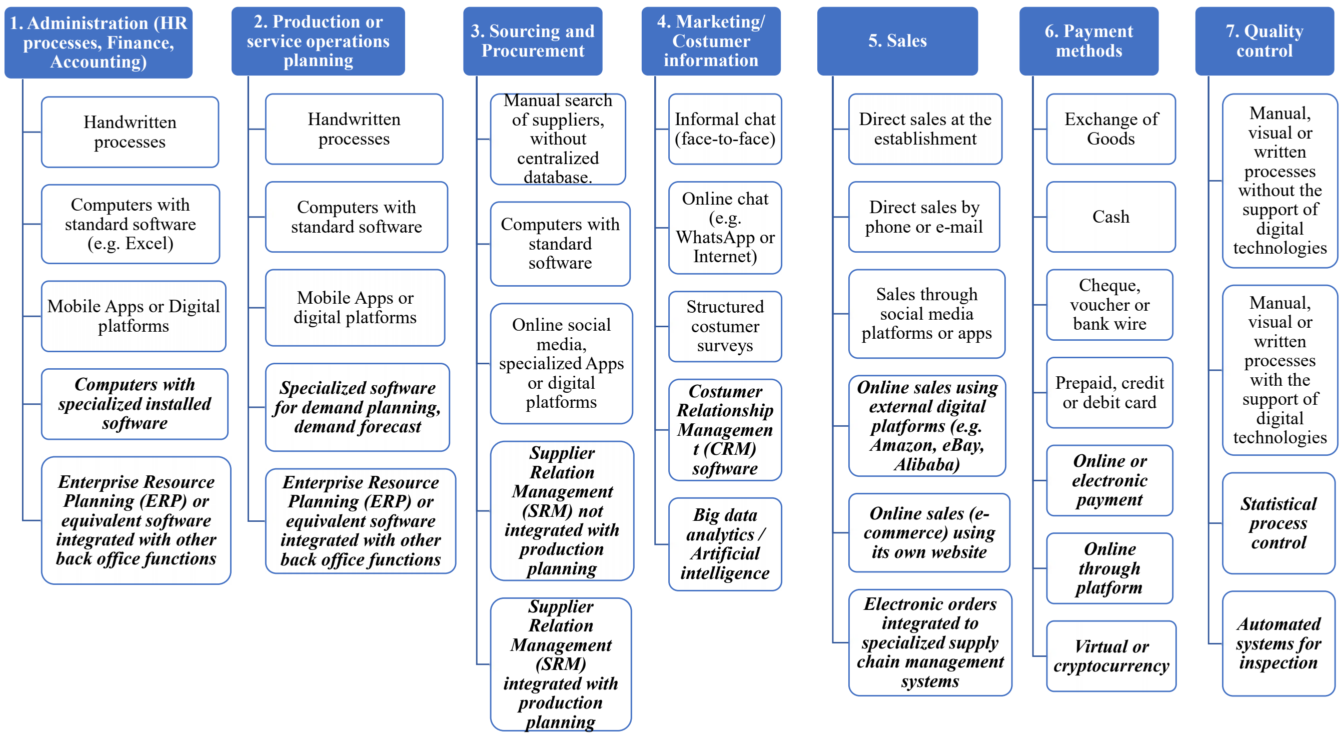

Figure 1: General Business Functions and Their Technologies

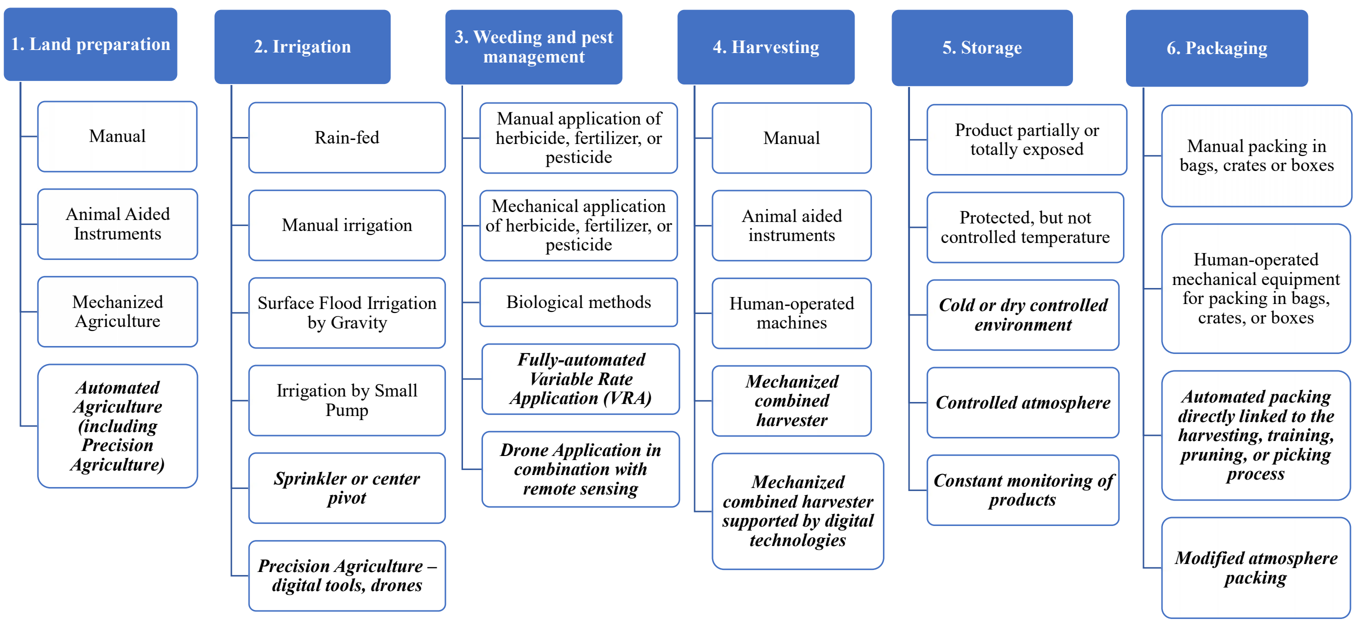

Figure 2: Sector Specific Business Functions and Technologies in Agriculture

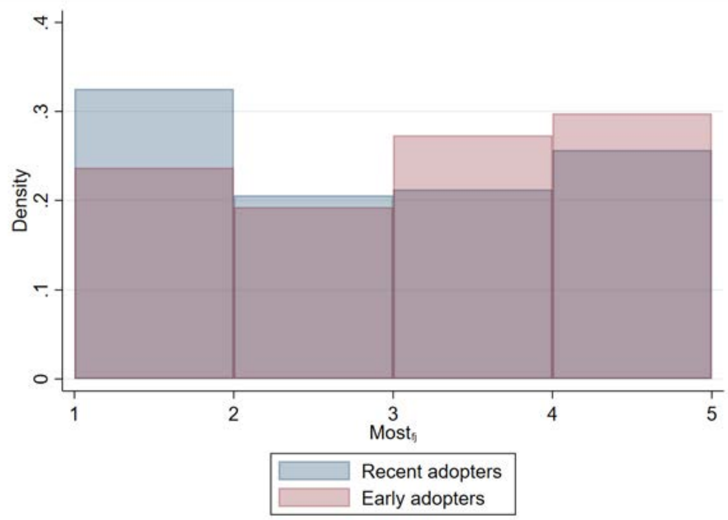

Figure 3: Distribution of MOST_{fj} conditional on adopting top-tiered technologies for early and late adopters

Note: Top-tier technologies are listed in Appendix A. Early and Recent adoption are defined at the BF level, and correspond to establishments that adopt the top-tier technologies earlier than, or after the median number of years since adoption, respectively.

Tables

Table 1: Number of establishments in FAT by country, sector and size

		Sector			Size		
	Total	Agri.	Manu.	Serv.	Small	Medium	Large
Bangladesh	903	-	744	159	361	232	310
Brazil*	1531	96	726	709	690	563	278
BurkinaFaso	600	80	142	378	335	187	78
Cambodia	794	-	333	461	583	142	68
Chile	1095	44	321	730	545	390	160
Croatia	710	46	272	392	472	183	55
Ethiopia	1476	149	747	580	999	330	147
Georgia	1800	196	768	836	741	632	427
Ghana	1262	85	350	827	774	382	106
India**	3242	101	1841	1300	1822	912	508
Kenya	1305	155	438	712	499	421	385
Korea	1551	128	658	765	656	569	326
Poland	1500	90	624	786	779	394	327
Senegal	1786	204	679	903	1219	395	172
Vietnam	1499	110	806	583	774	426	299
Total	21055	1485	9449	10121	11249	6158	3646

Note : * Brazil refers to state of Ceará; ** States of Tamil Nadu, Uttar Pradesh, Gujarat, and Maharashtra in India. The survey does not cover agriculture or services in Bangladesh, nor agriculture in Cambodia. In India, only the states of Gujarat and Maharashtra have agriculture included in the survey.

Table 2: Average level of technology measures

	$ANUM_{f,j}$	$NUM_{f,j}$	$MAX_{f,j}$	$MOST_{f,j}$	N_f
ABFs	2.1	2.1	2.8	2.0	4.8
GBFs	2.1	2.0	2.7	2.0	5.3
SSBFs	2.1	2.1	2.8	2.0	4.8

Notes : See Section 3.1 for definitions of variables. The table reports the average across the specific class of business functions, after averaging across establishments using sampling weights.

Table 3: Distributions of number of technologies, and $MAX_{f,j}$ in functions with only one technology

	$ANUM_{f,j}$					
	1	2	3	4	≥ 5	> 1
Percentage	37.4%	34.3%	17.1%	7.5%	3.7%	62.6%
	Conditional on $ANUM_{f,j} = 1$, $MAX_{f,j}$					
	1	(1,2]	(2,3]	(3,4]	(4,5]	
Percentage	52.5%	26.3%	10.5%	7.2%	3.4%	

Notes - The top panel of this table reports the distribution of $ANUM_{f,j}$. The bottom panel reports the distribution of $MAX_{f,j}$, conditional on the establishment using exactly one technology in the BF (i.e. $ANUM_{f,j} = 1$). The statistics are calculated using establishment-level sampling weights.

Table 4: Percentage of establishments with sophistication gaps

	GAP
ABFs	25%
GBFs	27%
SBFs	17%

Notes - Average across the specific class of business functions, after averaging across establishments using sampling weights.

Table 5: Relationship between technology measures

	$MOST_{f,j}$	$MAX_{f,j}$	$MOST_{f,j}$
$MAX_{f,j}$	0.55***		
	(0.01)		
$NUM_{f,j}$		0.85***	0.25***
		(0.01)	(0.01)
N	186503	186503	186503
R-squared	0.66	0.75	0.51
BF FE	Y	Y	Y
Firm FE	Y	Y	Y
Variation Explained	0.35	0.46	0.05

Notes - Estimates are from specifications 4, 5, and 6. To compute the last row, we first residualize the dependent and independent variables by regressing them on the fixed effects, and then we regress the residuals of the dependent on those of the independent. The reported figure is the corresponding R^2. Regressions are estimated using establishment-level sampling weights. Standard errors are clustered at the establishment level. *, ** and *** denote 10%, 5% and 1% significance respectively. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table 6: $MAX_{f,j} > MOST_{f,j}$ conditional on $NUM_{f,j} > 1$

		$MAX_{f,j}$ in			
	Overall	[1,2]	(2,3]	(3,4]	(4,5]
$\Pr(D_{fj} = 1 \mid NUM_{fj} > 1)$	62.6%	53.7%	67.0%	60.4%	71.1%

Notes - This table reports the probability of $D_{f,j} = 1$ (i.e. $MAX_{f,j} > MOST_{f,j}$), conditional on the establishment using more than one technology in the BF ($NUM_{f,j} > 1$). Columns 2-5 additionally condition on the value of $MAX_{f,j}$. Calculations made using establishment-level sampling weights

Table 7: Cross-establishment drivers of fraction of functions with $MAX_{f,j} > MOST_{f,j}$

	% functions with $MAX_{f,j} > MOST_{f,j}$
Size : Medium	-0.057***
	(0.006)
Size : Large	-0.038***
	(0.011)
Age: 6 - 10 Years	-0.014**
	(0.007)
Age : 11 - 15 Years	-0.055***
	(0.007)
Age : 16+ years	-0.026***
	(0.006)
Multi-establishment	0.006
	(0.006)
Foreign Owned	0.017**
	(0.008)
Exporter	-0.060***
	(0.007)
Constant	0.621***
	(0.033)
N	18322
R-squared	0.107
2-Dig. Sector FE	Yes
Country FE	Yes

Notes - Dependent variable is defined at establishment level as number of BFs with $MAX_{fj} > MOST_{fj}$, over number of BFs with $NUM_{f,j} > 1$. The base categories are Size: Small, and Age $\leq$ 5 Years. Establishments weighted by sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table 8: Average frequency of SP_j, IH_j, O_j and IN_j

		Relevant		
				Multi-establishment
Sector	Specialization	Inhouse	Outsourcing	Insourcing
Agriculture	0.20	0.95	0.05	0.06
Livestock	0.18	0.89	0.10	0.06
Food Processing	0.12	0.97	0.02	0.01
Apparel	0.15	0.91	0.08	0.06
Motor vehicles	0.66	0.64	0.31	0.20
Pharmaceuticals	0.32	0.94	0.04	0.06
Leather	0.41	0.68	0.32	0.07
Wholesale or retail	0.19	0.93	0.04	0.14
Financial services	0.12	0.94	0.04	0.03
Land transport	0.12	0.93	0.06	0.08
Health services	0.20	0.83	0.14	0.20
Overall	0.24	0.87	0.11	0.09

Notes - See definitions in Section 4. Averages are computed using sampling weights.

Table 9: Drivers of SP_j, O_j, and IN_j

	SP_j		O_j	IN_j
	(1)	(2)	(3)	(4)
$\overline{S}_j$	0.08***	0.14***	0.16***	0.33***
	(0.02)	(0.03)	(0.02)	(0.04)
$\overline{S}_j^2$	-0.02***	-0.04***	-0.03***	-0.06***
	(0.00)	(0.01)	(0.00)	(0.01)
$MAX_j - MOST_j$	-0.04***	-0.04***	-0.02***	-0.03***
	(0.01)	(0.01)	(0.00)	(0.01)
Age : 6-10 Years	-0.04***	-0.06***	-0.02***	0.07***
	(0.01)	(0.01)	(0.00)	(0.01)
Age : 11-15 Years	-0.02**	-0.02**	-0.02***	0.09***
	(0.01)	(0.01)	(0.00)	(0.01)
Age : 16+ Years	-0.02***	-0.04***	-0.04***	-0.01
	(0.01)	(0.01)	(0.00)	(0.01)
Size : Medium	-0.02***	-0.01*	-0.01***	-0.02**
	(0.01)	(0.01)	(0.00)	(0.01)
Size : Large	-0.00	0.03*	-0.02***	-0.00
	(0.01)	(0.02)	(0.01)	(0.02)
Multi-establishment	0.04***	0.01	-0.00	
	(0.01)	(0.01)	(0.00)	
Export	-0.04***	-0.04***	-0.01**	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Foreign owned	0.01	-0.00	0.02***	-0.03**
	(0.01)	(0.01)	(0.01)	(0.01)
% of main product in sales		0.06***		
		(0.01)		
N	11436	6355	11214	2319
R-squared	0.16	0.18	0.14	0.20
2-Dig. Sector FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes

Notes: Estimates from Specification 7. The variables SP_j, O_j , and IN_j are defined as in Section 4. The base categories are Size: Small, and Age: $\leq$ 5 Years. Establishments weighted by sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table 10: Statistics on distributions of MAX_j and $MOST_j$

	MAX_j					$MOST_j$				
Sector	Mean	SD	p10	p50	p90	Mean	SD	p10	p50	p90
Overall	2.64	0.78	1.67	2.56	3.75	2.03	0.65	1.24	1.95	2.93
Agriculture	2.68	0.91	1.56	2.59	3.90	2.10	0.69	1.21	2.05	3.05
Manufacturing	2.64	0.71	1.83	2.59	3.61	2.03	0.64	1.23	1.94	2.95
Services	2.64	0.80	1.61	2.54	3.80	2.03	0.65	1.24	1.95	2.93

Notes - Statistics are calculated using establishment-level sampling weights.

Table 11: Technology sophistication across countries

		MAX_j				$MOST_j$			
Country	$lnGDP$	All	Agri.	Manu.	Serv.	All	Agri	Manu	Serv
Korea	10.65	2.60	2.94	2.59	2.60	2.29	2.63	2.26	2.29
Poland	10.38	2.70	2.81	2.81	2.67	2.11	2.39	2.20	2.08
Croatia	10.17	3.21	3.10	3.10	3.25	2.33	2.40	2.15	2.39
Chile	9.78	2.74	2.76	2.66	2.76	2.22	2.32	2.14	2.24
Brazil	9.59	3.25	3.88	3.04	3.30	2.24	2.83	2.13	2.26
Georgia	9.46	2.49	2.46	2.55	2.48	2.05	2.18	2.12	2.03
Vietnam	8.92	2.53	2.57	2.48	2.55	1.92	2.04	1.86	1.93
India	8.81	2.72	2.23	2.67	2.76	2.00	1.84	1.93	2.04
Ghana	8.58	2.45	2.29	2.16	2.56	1.61	1.89	1.53	1.63
Bangladesh	8.44	2.16		2.16	1.77	1.73		1.73	1.39
Kenya	8.35	2.68	2.83	2.89	2.67	1.69	2.16	1.86	1.67
Cambodia	8.09	1.86	1.21	1.89	1.86	1.57	1.10	1.65	1.56
Senegal	8.09	1.82	1.60	1.72	1.96	1.32	1.29	1.27	1.39
Ethiopia	7.70	1.64	1.91	1.85	1.63	1.37	1.61	1.50	1.37
BurkinaFaso	7.66	1.84	1.98	1.82	1.85	1.25	1.31	1.31	1.26
$Corr$		0.77	0.71	0.76	0.74	0.93	0.85	0.91	0.92
SD	0.97	0.49	0.68	0.46	0.52	0.37	0.52	0.33	0.39
Cov		0.37	0.48	0.34	0.37	0.33	0.44	0.29	0.34

Notes: Country marginal effects of MAX_j and $MOST_j$ estimated by regressing MAX_j and $MOST_j$ on country and 1-Dig. sector FE, using establishment-level sampling weights. $lnGDP$ is the log per-capita GDP from Penn World Tables (2019). SD is the standard deviation of the column, $Corr$ and Cov denote the correlation and covariance of the column with $lnGDP$.

Table 12: Technological sophistication and establishment characteristics

	(1) MAX_j	(2) $MOST_j$
Size: Medium	0.31***	0.23***
	(0.01)	(0.01)
Size: Large	0.67***	0.51***
	(0.02)	(0.02)
Age: 6 to 10	0.09***	0.14***
	(0.01)	(0.01)
Age: 11 to 15	0.06***	0.15***
	(0.01)	(0.01)
Age: 16+	0.03**	0.05***
	(0.01)	(0.01)
Foreign owned	0.32***	0.28***
	(0.02)	(0.01)
Exporter	0.26***	0.21***
	(0.02)	(0.01)
Multi-establishment	0.34***	0.23***
	(0.01)	(0.01)
N	19253	19253
R-squared	0.37	0.31
2-Dig. Sector FE	Yes	Yes
Country FE	Yes	Yes

Notes - Estimates of MAX_j and $MOST_j$ on establishment characteristics using establishment-level sampling weights. The base categories are Size: Small, and Age: $\leq$ 5 Years. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table 13: Technology sophistication vs. narrow establishment-level technology measures

	$\overline{S}_j$								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Computers	0.89***								
	(0.01)								
ERP		0.94***							
		(0.01)							
Robots			0.97***						
			(0.04)						
$\overline{S}_{BusAdmin,j}$				0.41***					
				(0.00)					
$\overline{S}_{Payments,j}$					0.38***				
					(0.01)				
$\overline{S}_{Fabrication,j}$						0.49***			
						(0.01)			
$\overline{S}_{GBF,j}$							0.94***		
							(0.00)		
$\overline{S}_{SSBF,j}$								0.15***	
								(0.00)	
$\overline{S}_{GBF,j}$ * Agriculture									0.64***
									(0.02)
$\overline{S}_{GBF,j}$ * Manufacturing									0.77***
									(0.00)
$\overline{S}_{GBF,j}$ * Services									0.95***
									(0.00)
$\overline{S}_{SSBF,j}$ * Agriculture									0.34***
									(0.01)
$\overline{S}_{SSBF,j}$ * Manufacturing									0.21***
									(0.00)
$\overline{S}_{SSBF,j}$ * Services									0.00***
									(0.00)
N	20407	19530	7238	20383	20566	6908	20995	18316	18310
R-squared	0.18	0.31	0.07	0.56	0.21	0.31	0.94	0.07	0.95
1-Dig. Sector FE	No	No	No	No	No	No	Yes	Yes	Yes

Notes : Estimates from regressing $\overline{S}_j$ on establishment-level technology measures. Computers, ERP, and Robots are binary variables. Computers is 1 if establishment has computers, electricity and internet, and 0 otherwise. ERP and robots are 1 if establishments have ERP and robots respectively, and are 0 otherwise. Column 3 is only run for manufacturing establishments. The independent variables in columns 4-6 reflect the technology sophistication (average of MAX_j and $MOST_j$) in specific business functions namely - business administration, payments, and fabrication (only for manufacturing establishments). Columns 7 and 8 cover average sophistication of GBFs and SSBFs respectively. Agriculture, Manufacturing and Services are binary variables that take the value of 1 if the establishment operates in the 1-digit sector and 0 otherwise. All regressions use establishment-level sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table 14: Productivity and technological sophistication

	log(sales per worker)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
K_j	0.234***	0.266***	0.223***	0.233***	0.231***	0.342***	0.442***	0.234***	0.281***	0.218***	0.244***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.023)	(0.024)	(0.008)	(0.009)	(0.011)	(0.012)
H_j	0.191***	0.585***	0.150***	0.189***	0.202***	0.507**	0.825***	0.164***	0.604***	0.165***	0.586***
	(0.040)	(0.042)	(0.040)	(0.040)	(0.040)	(0.227)	(0.236)	(0.061)	(0.063)	(0.058)	(0.060)
$\overline{S}_j$	0.493***	0.631***	0.504***	0.460***	0.422***	0.648***	1.023***	0.584***	0.701***	0.458***	0.587***
	(0.019)	(0.020)	(0.019)	(0.020)	(0.022)	(0.088)	(0.086)	(0.023)	(0.025)	(0.030)	(0.030)
Management (Z-Score)				0.062***	-0.003						
				(0.011)	(0.018)						
$\overline{S}_j$ * D(High Management)					0.068***						
					(0.015)						
Constant	6.118***	5.951***	7.948***	6.206***	6.217***	5.402***	3.067***	6.471***	6.326***	7.502***	7.410***
	(0.175)	(0.147)	(0.310)	(0.175)	(0.175)	(0.419)	(0.233)	(0.124)	(0.108)	(0.420)	(0.139)
N	13046	13046	13046	13038	13038	825	825	6032	6032	6189	6189
R-squared	0.407	0.234	0.435	0.409	0.410	0.716	0.577	0.480	0.327	0.382	0.186
2 Dig. Sector FE	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	No		Yes	Yes	Yes	No	Yes	No	Yes	No
2 Dig. Sector X Country FE			Yes								
Data	All	All	All	All	All	Agri.	Agri.	Manu.	Manu.	Serv.	Serv.

Notes : Estimates of specification (8). D(High Management) is a dummy that takes the value 1 if the management z-score of the establishment defined in Section 2 is above the median, and 0 otherwise. All regressions estimated using establishment-level sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table 15: Productivity and dimensions of technology sophistication

	log(sales per worker)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
K_j	0.235***	0.267***	0.254***	0.287***	0.264***	0.458***	0.191***	0.222***	0.292***	0.340***
	(0.007)	(0.007)	(0.008)	(0.009)	(0.031)	(0.033)	(0.008)	(0.009)	(0.017)	(0.018)
H_j	0.233***	0.633***	0.119**	0.482***	0.126	0.151	0.152*	0.677***	0.018	0.417***
	(0.041)	(0.042)	(0.053)	(0.054)	(0.236)	(0.247)	(0.079)	(0.079)	(0.086)	(0.086)
MAX_j	0.084***	-0.037								
	(0.023)	(0.023)								
$MOST_j$	0.433***	0.740***								
	(0.025)	(0.027)								
$\overline{S}_{GBF,j}$			0.304***	0.452***	0.870***	1.142***	0.418***	0.629***	0.224***	0.315***
			(0.028)	(0.029)	(0.089)	(0.098)	(0.032)	(0.034)	(0.050)	(0.050)
$\overline{S}_{SSBF,j}$			0.081***	0.170***	0.056	0.067	0.110***	0.043	0.094**	0.265***
			(0.024)	(0.026)	(0.071)	(0.079)	(0.027)	(0.030)	(0.042)	(0.047)
Constant	6.103***	5.971***	5.958***	5.594***	5.165***	2.794***	7.015***	7.024***	7.056***	6.432***
	(0.174)	(0.145)	(0.184)	(0.156)	(0.459)	(0.293)	(0.126)	(0.115)	(0.220)	(0.207)
N	13046	13046	8877	8877	784	784	4888	4888	3205	3205
R-squared	0.410	0.250	0.407	0.254	0.660	0.497	0.460	0.299	0.376	0.211
2 Dig. Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Data	All	All	All	All	Agri.	Agri.	Manu.	Manu.	Serv.	Serv.

Notes : All regressors are establishment-level measures. All regressions estimated using establishment-level sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table 16: Technology and productivity: non-linearities and appropriateness

	log(sales per worker)							
	Non-Linearities		Appropriateness					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K_j	0.23***	0.24***	0.23***	0.11***	0.31***	0.24***	0.24***	0.23***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
H_j	0.19***	0.24***	0.20***	0.45***	0.09**	0.25***	0.67***	0.10*
	(0.04)	(0.04)	(0.04)	(0.09)	(0.05)	(0.04)	(0.07)	(0.05)
$\overline{S}_j$	0.56***		0.51***	0.47***	0.50***		0.34***	0.54***
	(0.04)		(0.02)	(0.04)	(0.02)		(0.03)	(0.03)
$\overline{S}_j$ * D(High Sophistication)	-0.03*							
	(0.02)							
D($1.5 \leq \overline{S}_j \leq 2.5$)		0.31***				0.31***		
		(0.04)				(0.04)		
D($2.5 \leq \overline{S}_j \leq 3.5$)		0.75***				0.78***		
		(0.05)				(0.05)		
D($\overline{S}_j \geq 3.5$)		1.11***				1.21***		
		(0.07)				(0.08)		
$\overline{S}_j$ * D(High Income)			-0.07				-0.13 **	-0.00
			(0.04)				(0.06)	(0.06)
D($2.5 \leq \overline{S}_j \leq 3.5$) * D(High Income)						-0.12**		
						(0.06)		
D($\overline{S}_j \geq 3.5$) * D(High Income)						-0.33***		
						(0.11)		
Constant	6.02***	6.77***	6.08***	11.34***	5.31***	6.77***	6.57***	6.03***
	(0.18)	(0.17)	(0.18)	(0.36)	(0.18)	(0.17)	(0.22)	(0.31)
N	13046	13046	13046	2104	10942	13046	6383	6663
R-squared	0.41	0.40	0.41	0.30	0.38	0.40	0.43	0.42
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Countries	All	All	All	High Income	Low Income	All	High Emp.	Low Emp.

Notes : D(.) are binary variables that take the value 1 if the establishment/country satisfies the condition in parenthesis and 0 otherwise. High Sophistication represents that the establishment has above-median $\overline{S}_j$; the different intervals for $\overline{S}_j$ represent that the establishment's $\overline{S}_j$ is in the given interval; high income is satisfied if the establishment is in one of the three high-income countries, which are - South Korea, Poland and Croatia. High Emp. and Low Emp. are categories defined on the basis of above and below median number of employees. Base category for the high-income countries in column 6 is D($\overline{S}_j < 2.5$)*D(High Income). The first two sophistication categories have been merged for high-income countries because only 49 establishments (1% of all high-income estab.) belong to the group D($\overline{S}_j < 1.5$). All regressions are estimated using establishment-level sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Contents (Appendix)

A The FAT survey and section 2 results

This section provides more details on the Firm Adoption of Technologies (FAT) survey and its implementation. We start with a description of the grid of technologies in FAT. Then we describe the sampling frameworks used and the construction of sampling weights. We finalize describing all the tests conducted to minimize potential biases, including validation exercises ex post implemented with with external data sources.

A.1 The survey

The FAT survey is a multi-country, multi-sector, representative firm-level survey. It collects information on the technologies used by firms in specific business functions that encompass the key activities that each firm conducts. Compared to existing firm-level surveys, the FAT survey covers a significantly larger number of technologies and business functions (Table A.1), and a wider range of sectors; for example, it covers agriculture distinguishing between crops and livestock.

Table A.1: Coverage of Firm-Level Technology Surveys

Surveys	# of Technologies	# of Business Functions	Includes Firms in Agriculture
Firm-level Adoption of Technology (FAT) Survey	302	63	Yes
Survey of Advanced Technology (SAT)	57	3	No
Community Survey on ICT Usage and E-Commerce in Enterprises	9	0	No
Information & Communication Technology Survey (ICTS)	4	0	No
Annual Business Survey (ABS) 2019	5	0	No

Note: The Number of technologies and business functions are computed by authors.

The FAT survey addresses important knowledge gaps compared to other surveys measuring technology at the firm or establishment level. For starters, the number of technologies covered is rather limited when compared to how many technologies are involved in production processes. Second, their focus on the presence of advanced technologies makes it impossible to understand how production takes place in companies without such advanced technologies. This concern is most relevant in developing countries where advanced technologies have diffused less. Third, since their unit of analysis is the firm, existing studies are not designed to analyze what business functions benefit from each technology. This drawback is particularly problematic for general technologies that can be relevant for multiple business functions. Finally, existing surveys largely omit questions about how intensively a technology is employed in the firm, and therefore, they do not reveal whether a technology that is present is widely utilized or just marginally.

Specifically, the FAT survey comprises five sections:

- Module A– Collects general information about the characteristics of the establishment; such as sector, multi-establishment and ownership.
- Module B – Covers the technologies used in seven general business functions.
- Module C – Covers the use of technologies for functions that are specific to each of 11 agriculture, manufacturing, and services sectors
- Module D – Includes questions about the drivers and barriers for technology adoption.
- Module E – Collects information on employment, balance sheet and performance, which allow us to compute labor productivity and other measures at the establishment level.

A.1.1 The Grid. Business functions and relevant technologies

We construct a technology grid that identifies the main business functions and the key technologies used to carry out the tasks of each business function. To design modules B and C, the survey draws upon the knowledge of experts in production and technology in various fields and sectors. These experts provided their knowledge on: i) what are the key general and sector-specific business functions, ii) what are the different technologies used to conduct the main tasks in each function, and iii) how are the different technologies related both in terms of their sophistication and the degree of substitutability between them.

First, we started with desk research revising the specialized literature identifying business functions and technologies across the value chain.[58] Second, for each sector, as well as for the general business functions, we hold meetings with private sector specialists at the World Bank Group to validate the initial findings and start to define the key business functions and technologies. Third, we hold meetings with Lead and Senior Economists across the World Bank Group, including the International Finance Corporation (IFC), from different fields of specialization and wide experience with sectoral projects in several countries (e.g. agriculture, manufacturing, retail, transport, health, etc.). Fourth, we hold meetings and validation exercises with external senior consultants, with wide experience on the field (e.g. at least 15 years), including experience with firms in developing countries as well as advanced economies.

The source of external senior consultants in the last layer of quality control varied across sector. For agriculture and livestock, the validation exercise was conducted with agricultural engineers and researchers from Embrapa, an agricultural research institution from Brazil. For food processing, wearing apparel, pharmaceutical, transport, and retail, as well as for

[58]This process involved the revision of peer-review journals and reports from international organizations and industry associations.

the general business functions, the team hired external consultants through a large management consultant organization. For automotive sector, the team has hired a senior consultant directly. For health, the team invited directly five physicians with different field of specializations and practical experience in hospitals in clinics in the United States and low income countries in Saharan Africa.

The validation exercise with sector specialists were organized as follows. First, the team would explain the purpose of the project, present the initial findings, and share a draft with identified business functions and technologies associated with them. The sector specialists would have between one and two weeks to reflect on the material to validate them or propose a new combination of business functions and technologies associated with them. After receiving the revised material, a second meeting with sector specialists would be organized with the FAT survey team to discuss the proposal and converge towards an updated combination of business's functions and technologies.

In what follows we describe the grids for both types of business functions.

A.1.2 General Business Functions

Figure 1 shows the 7 general business functions in FAT and the possible technologies used to conduct them. The business functions identified are: business administration (HR processes, finance,..), production planning, procurement and supply chain management, marketing and product development, sales, payment methods, and quality control. These are business functions that in addition to being central in the functioning of the firm, are also retained in some capacity (or some tasks) within the firm. The technologies used for these business functions tend to be more available and off-the-shelf technologies, often ICT technologies. For example, for administrative processes, these range from handwritten processes (the least sophisticated) to the use of enterprise resource planning which are software that allow for real time, integrated management of the main business processes. With the help of management consultants, we identify the technologies feasible for each business function and develop similar rankings of sophistication based on the consultants understanding of the number of tasks and complexity that the technologies can handle.

One important characteristic of the grid is that the sophistication rankings are not fully hierarchical for all business functions. In the case of sales, for example, firms can use various technologies, and while online sales are more sophisticated technologies that sales on the phone or email, there is no clear sophistication ranking between sales made in the company's website or using online platforms; both are complementary. A similar example occurs with payment methods; firms may use a variety of them, often depending on the financial infrastructure in the country.

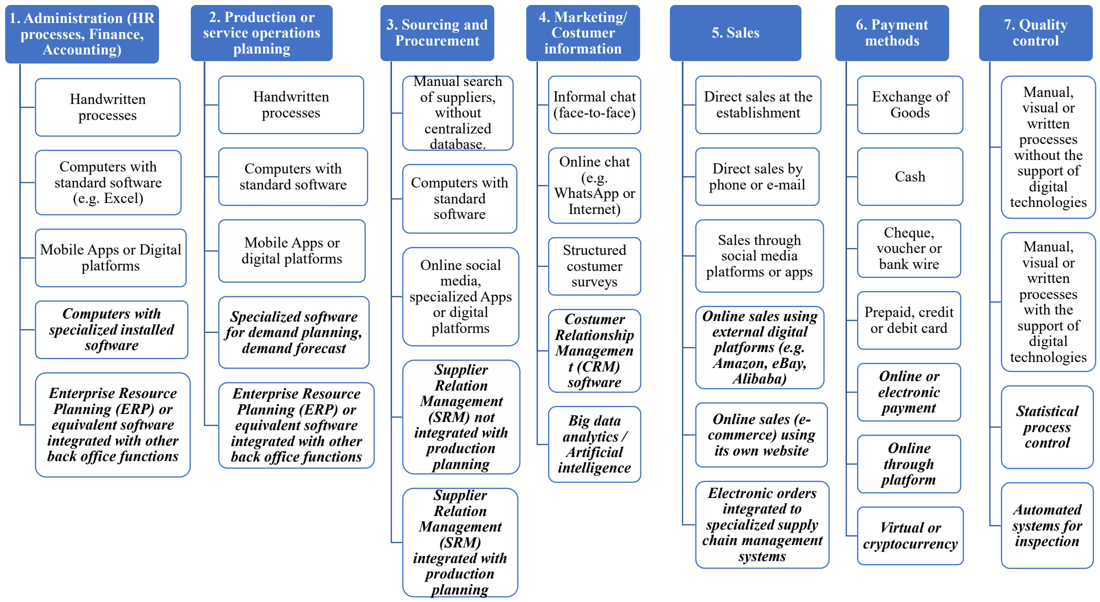

Figure A.1: General Business Functions and Their Technologies

A key advantage of the grid structure is that it allows to accommodate the use of more than one technology by business function. The survey questionnaire is implemented so respondents are asked first about the use of each of the technologies in the grid. Then, for those technologies selected in each business function, the respondent is asked to identify the one that is more intensively used in implementing the tasks of the business function. Finally, when using one of the most advanced technologies, the respondent is also asked to provide the year of adoption. This allows to uncover new facts about technology adoption and use by allowing to build new measures of technology sophistication at the business function level based on extensive measures, the most sophisticated technology, and intensive measures, the technology used more intensively. It, also allows to calculate measures of diffusion lags for advanced technologies.

A.1.3 Sector Specific Business Functions

For the sector-specific technologies, a similar approach was used to identify key business functions and associated technologies in 12 sectors of activity across agriculture, manufacturing, and services (including agriculture-crops; livestock; food processing; wearing apparel; leather and footwear; automotive; pharmaceuticals; wholesale and retail; transportation; financial services; health services; other manufacturing). One business function, fabrication, was also included for all manufacturing sectors. The identification of key business functions and the frontier in each sector required a significant interaction with several sector specialists. These functions tend to be associated with sector-specific production processes.

Here, we present all sector specific business functions and associated technologies covered by the FAT survey in the first and second phase of data collection. These figures complement the information provided in Section 2, particularly Figure 1 and Figure A.4, which describe the functions and associated technologies for GBFs and food processing, among SSBFs. The complementary information is provided for all SSBFs (Agriculture - Crops (Figure A.2), Livestock (Figure A.3), Food Processing (Figure A.4), Wearing Apparel (Figure A.5), Leather and Footwear (Figure A.6), Automotive (Figure A.7), Pharmaceutical (Figure A.8), Wholesale and Retail (Figure A.9), Transportation (Figure A.10), Financial Services (Figure A.11), Health Services (Figure A.12), and Other Manufacturing (Figure A.13)). [59]

[59] As the survey is rolled out in other countries, the number of additional sectors included in the survey is also increasing.

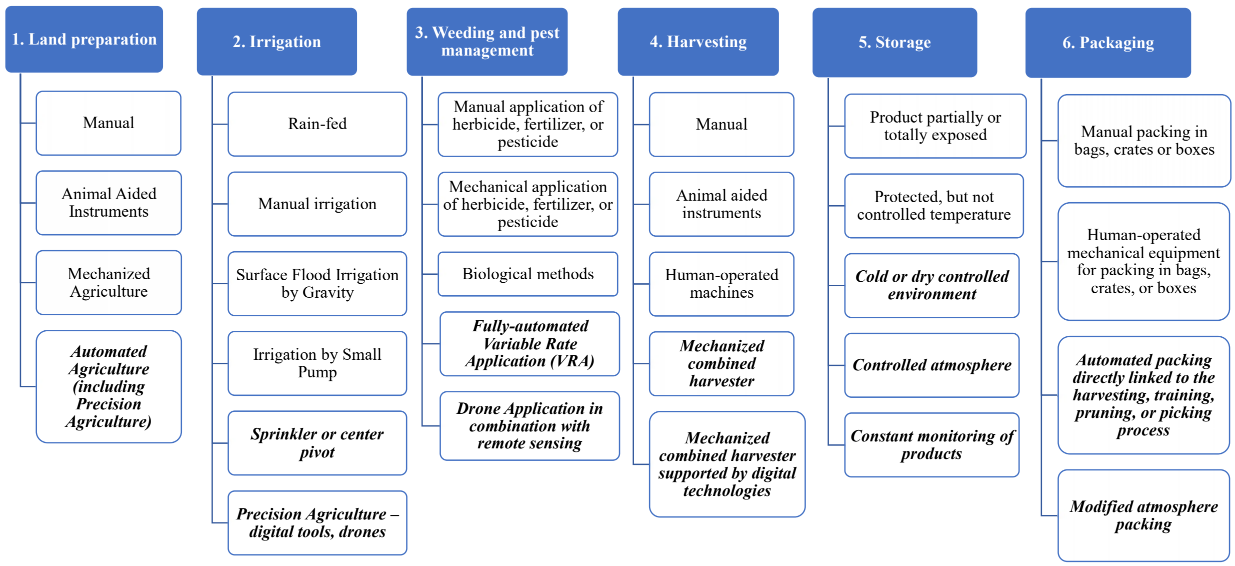

Figure A.2: Sector Specific Business Functions and Technologies in Agriculture - Crops

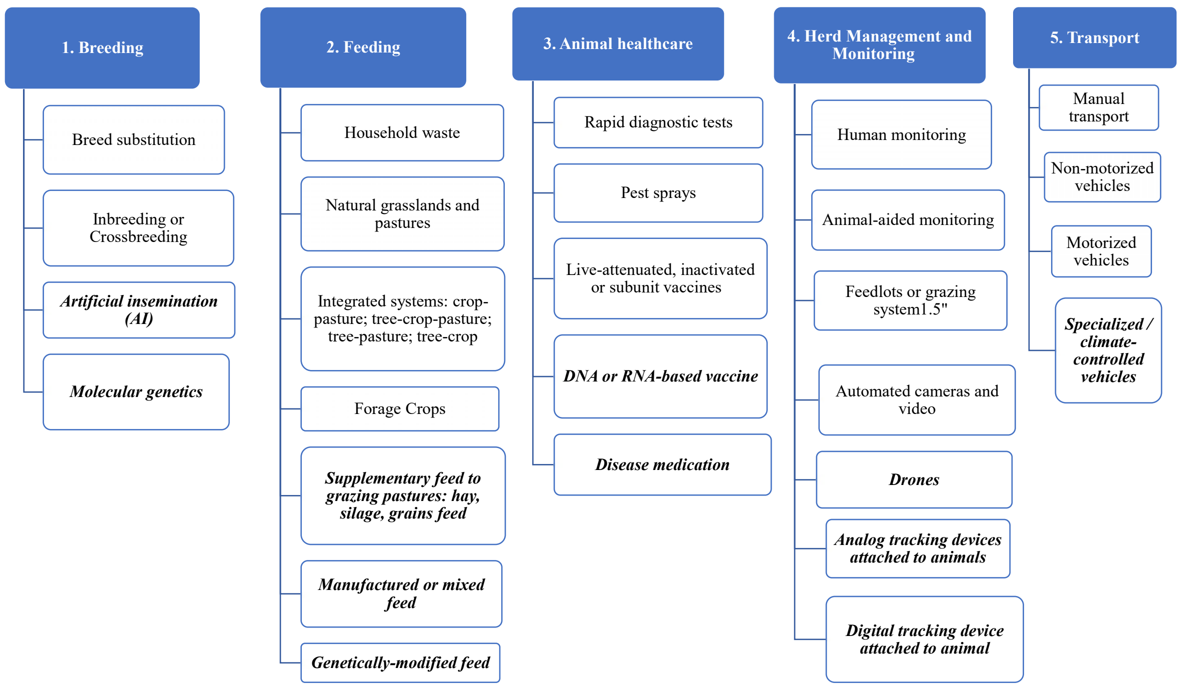

Figure A.3: Agriculture - Livestock: Business Functions and Technologies

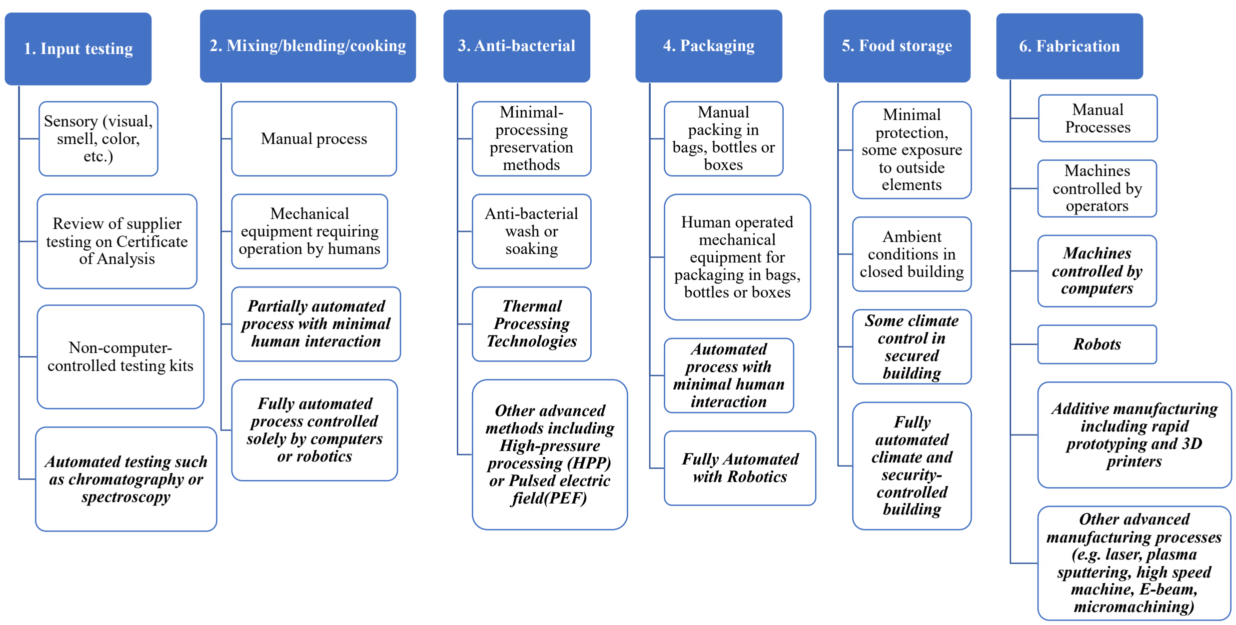

Figure A.4: Food Processing: Business Functions and Technologies

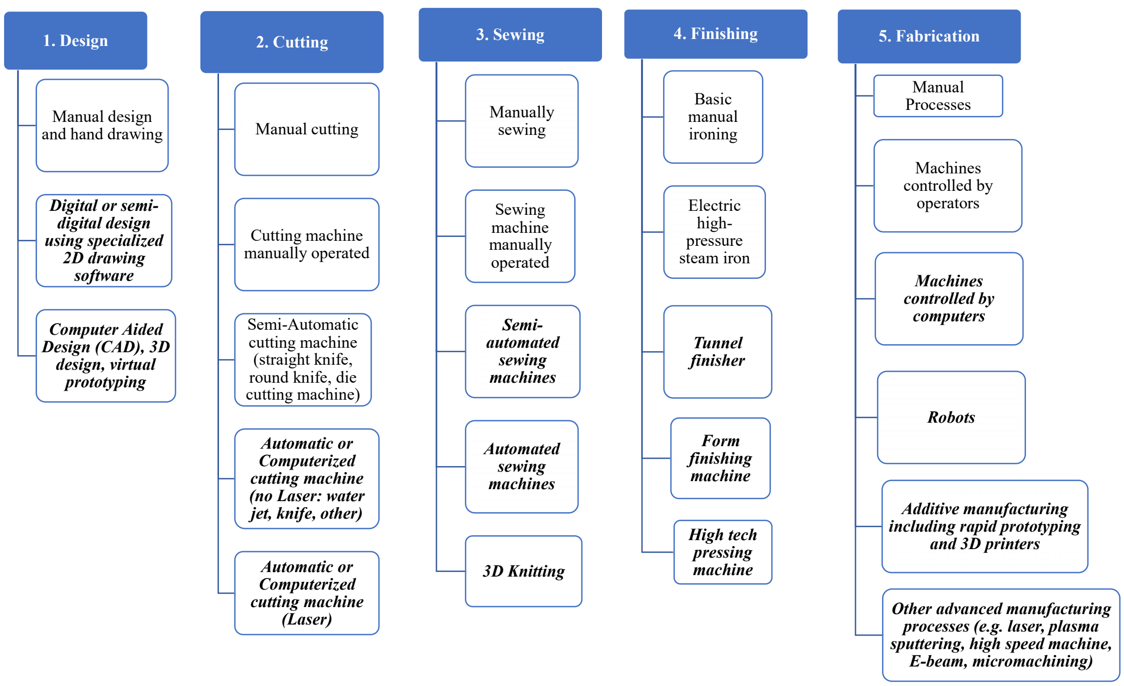

Figure A.5: Wearing Apparel: Business Functions and Technologies

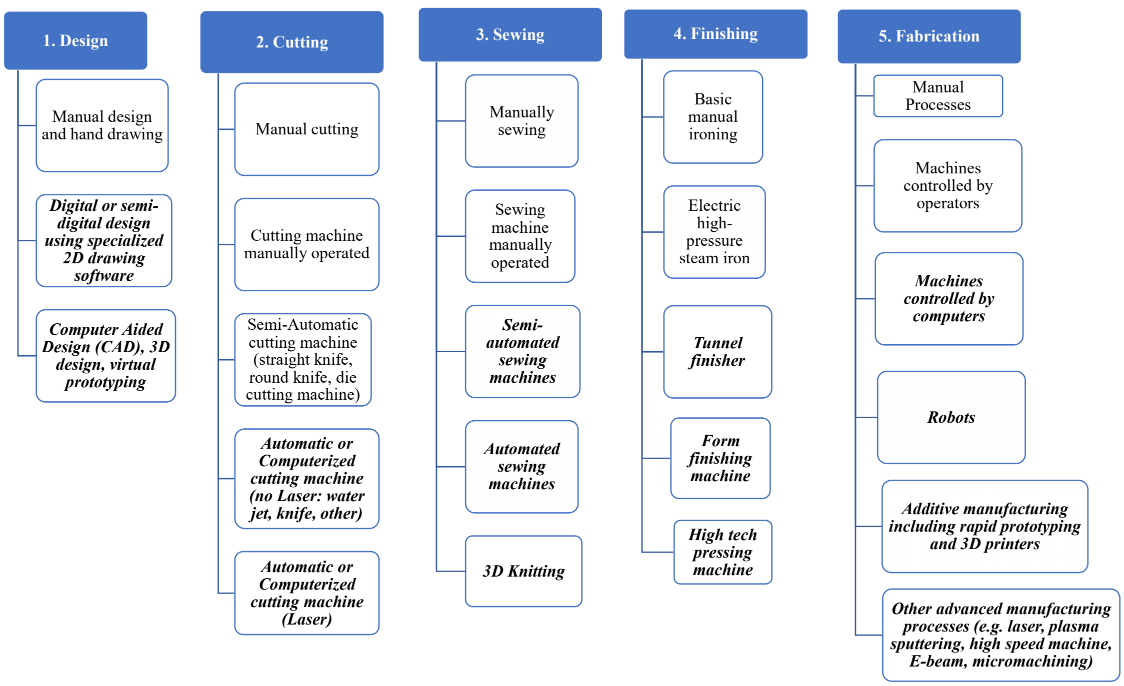

Figure A.6: Leather and Footwear: Business Functions and Technologies

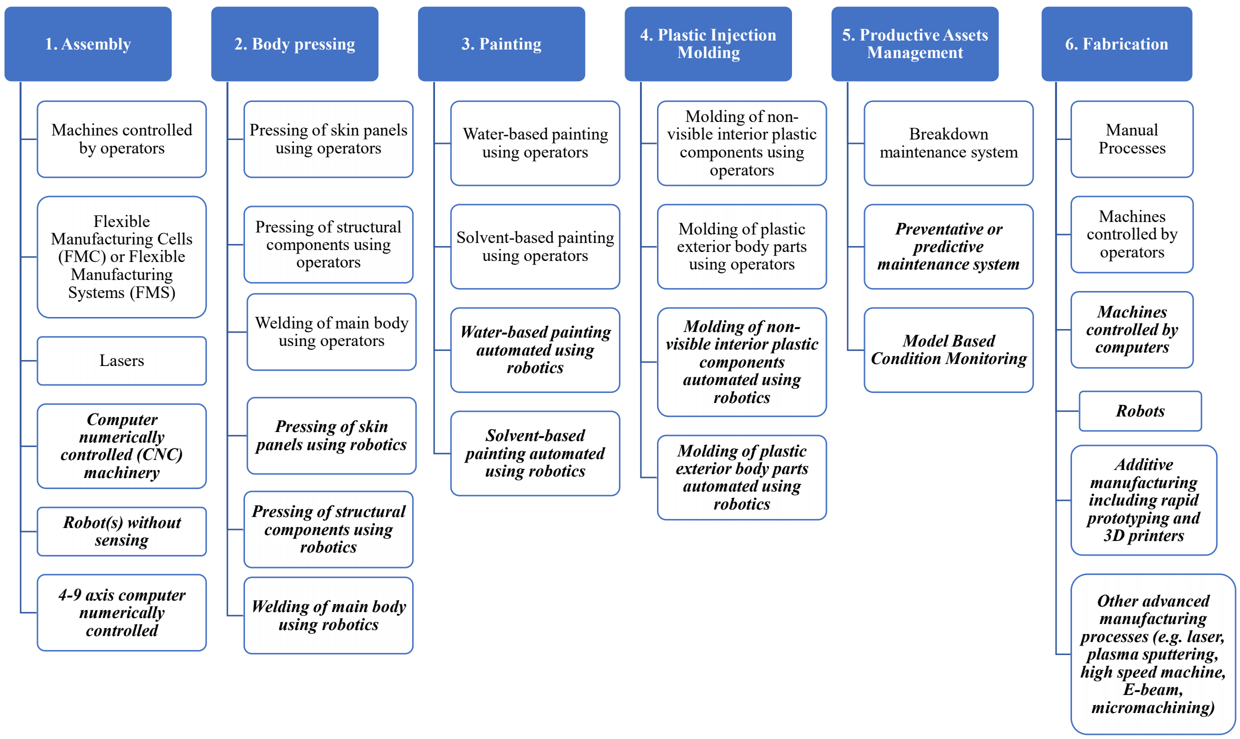

Figure A.7: Automotive: Business Functions and Technologies

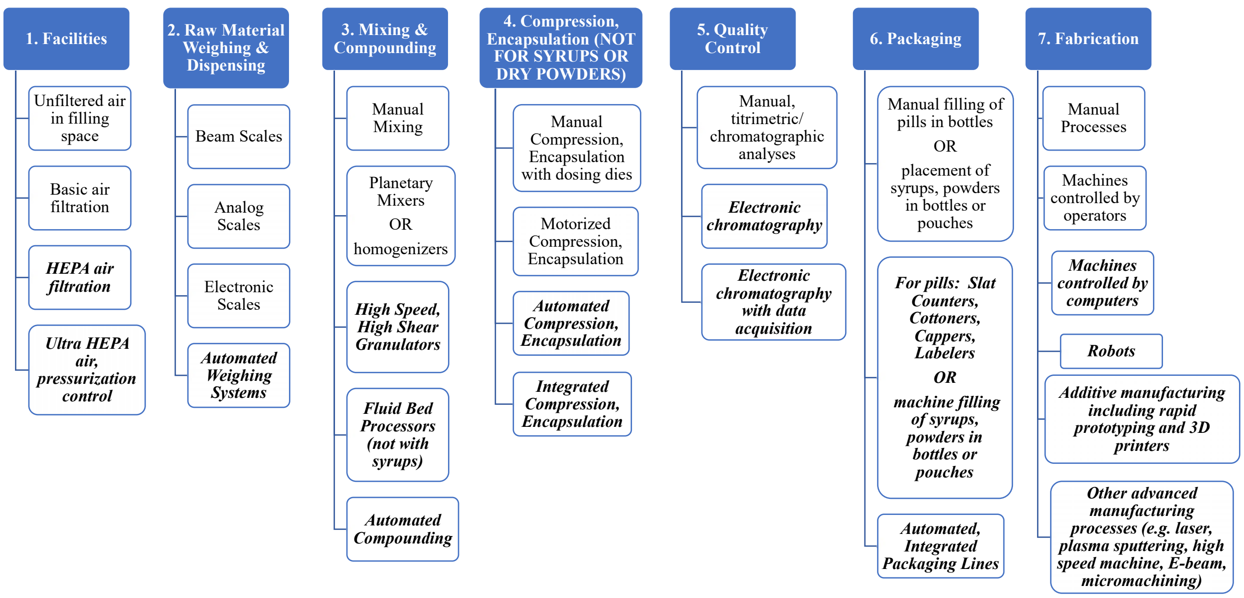

Figure A.8: Pharmaceutical: Business Functions and Technologies

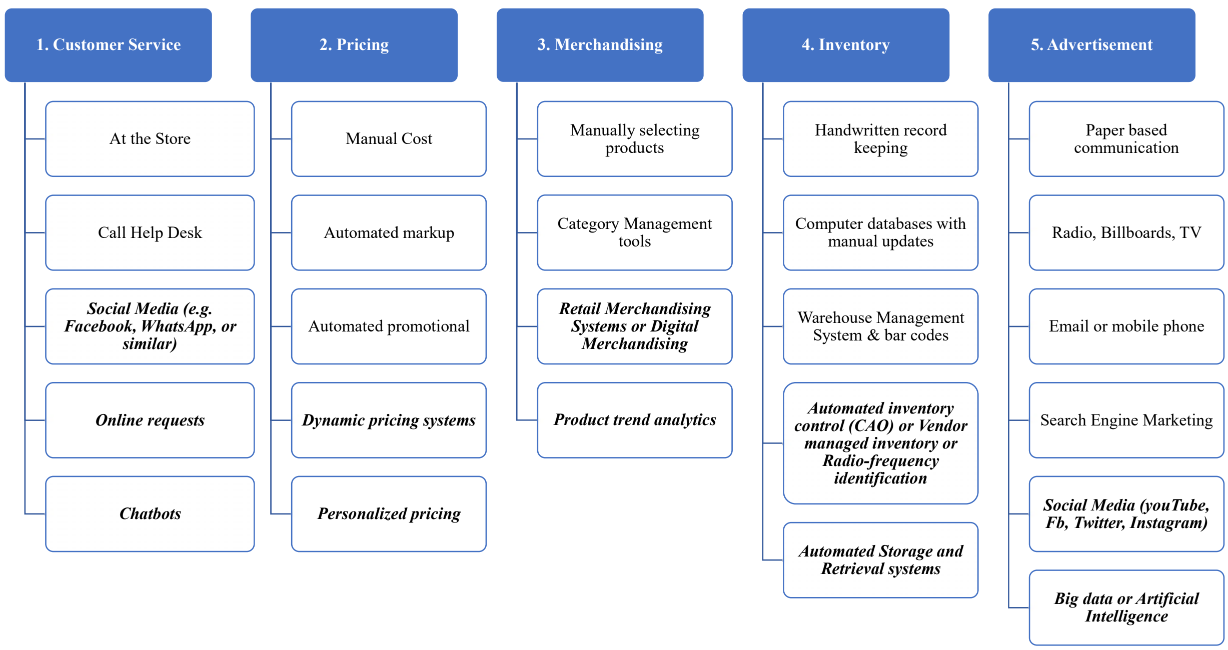

Figure A.9: Wholesale and Retail: Business Functions and Technologies

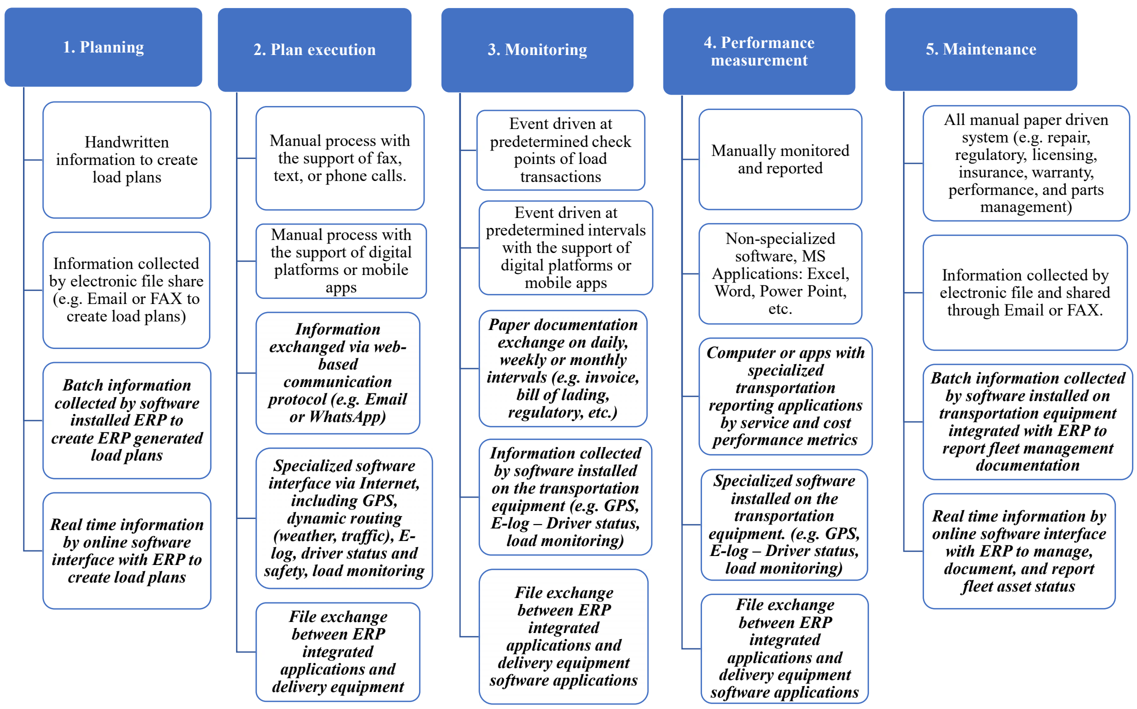

Figure A.10: Land Transportation: Business Functions and Technologies

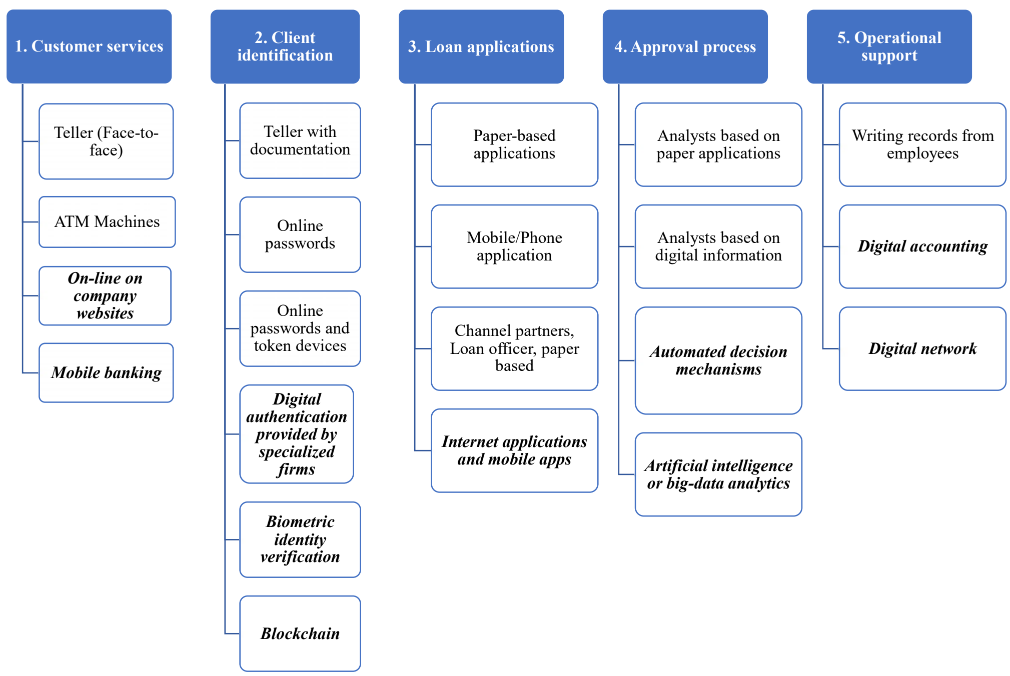

Figure A.11: Financial Services: Business Functions and Technologies

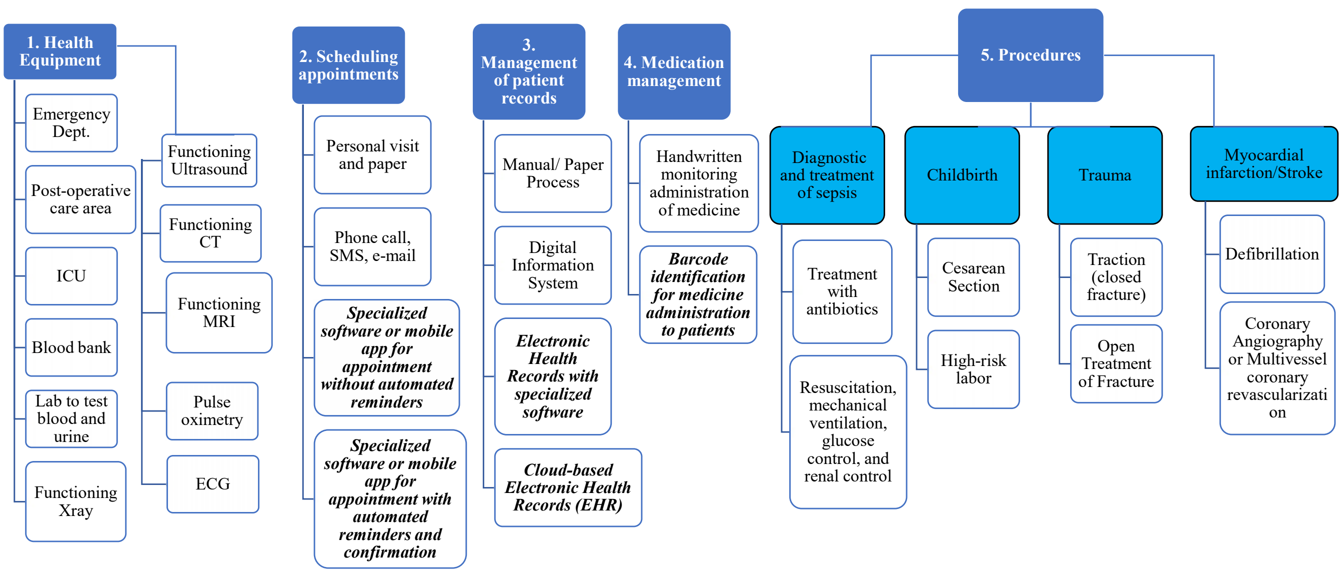

Figure A.12: Health Services: Business Functions and Technologies

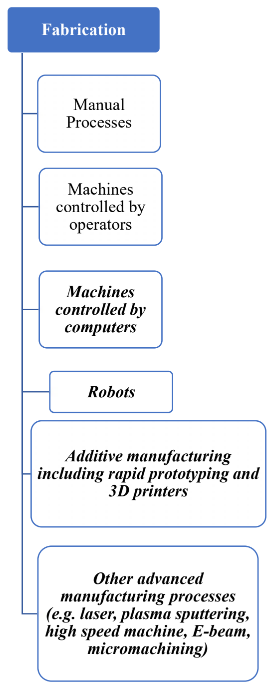

Figure A.13: Other Manufacturing: Business Functions and Technologies

For sector-specific business functions, digital technologies tend to be embedded in other technologies that are usually at the frontier. This is a common feature, particularly in agriculture and manufacturing, and has important implications in terms of the costs of adoption and the importance of network effects. For example, among methods commonly used by agricultural firms to perform harvesting, the most basic option is to harvest manually, followed by animal-aided instruments, human-operated machines, or a single tractor with one specific function (such as a single-axle tractor), a combined harvester (machines or tractors that combine multiple functions fully operated by the worker), and combined harvester using the support of digital technologies (such as global positioning systems [GPS] or computing systems integrated with the tractor). Unlike GBFs, the application of digital technologies for harvesting requires other sophisticated equipment or machines.

In addition to the possibility of computing different measures of technology sophistication for sector specific business functions, an important feature of the sector specific grid is the fact that it includes screening questions that allow for the fact that not all the business functions are carried out within the establishment. In other words, not all entries in the grid need to be implemented at the establishment or at the firm. While the tasks of most general business functions related to management and organization are usually carried out within the boundaries of the firm - either in the same establishment or in another establishment of the firm if multi establishment - some sector specific business functions can be carried out in another establishment within the same firm (insourcing), or they can be (outsourced) to a different firm. Our approach is, therefore, rooted in a view of the firm similar to Coase (1937), where firms are agents coordinating and implementing tasks. The advantage of this approach is twofold. In addition to the fact that this approach allows a better identification of technology and its use as described above, it allows to study critical questions such as the organization of the firm and tasks (Williamson, 1979), and more importantly the relationship between organizational modes, transaction costs and technological choice (Williamson, 1988).

After finalizing the FAT questionnaire, we pre-piloted it in Brazil and Senegal. We personally conducted the face-to-face interviews, in collaboration with enumerators and supervisors trained to conduct data collection with firms from different sectors and size groups. In the pre-pilot stage, we tested if the business functions and technologies covered by the questionnaire were comprehensive and clearly understood by respondents, through detailed discussions and follow up questions with representative of firms, which led us to make the necessary adjustments to the survey. For example, we experimented with survey designs that asked about the fraction of time/output/processes that were conducted with each of the technologies in the business function. We decided against using this approach to reflect the intensity of use of technologies because it was harder for respondents to answer precisely, and as a result led to a more subjective interpretation, which made the comparability of answers across business functions and companies harder to interpret.

A.1.4 Barriers and Drivers

In addition to the information on the technologies used by firms, the survey also collects information on potential drivers of and barriers to technology adoption. First, the survey asks whether the firm acquired new machines, equipment, and software in the last three years; and in the case of machines or equipment, whether these were leased, purchased as new or secondhand. The survey also asks questions on links to larger firms and multinationals, either via value chain linkages as a supplier or buyer, or via the CEO previous experience working in a MNE or a large firm.

The survey also asks questions about access to finance and trade status. The first question is about having secured a loan in the previous three years for purchasing equipment, machinery or software. On more general access to finance, the survey asks how many times the establishment needed to borrow money to expand production but could not obtain finance. On trading status, the survey asks whether the firm is an importer, an exporter; and if an exporter, what is the share of sales that is exported.

A key complementary factor for technology adoption is the quality of management. The survey pays special attention to management by collecting information on the top manager's background and on management practices. Specifically, FAT asks about the level of education attainment of the top manager in the establishment, whether she has studied abroad, and whether she has experience in multinationals. In addition, the survey contains four questions about management practices. These include four questions from MOPS (Bloom et al., 2016) on the number of KPIs, the frequency with which they are monitored, the horizon of production targets and a question on the use of formal incentives. Though the information we collect on management practices is more restricted than the sixteen questions in MOPS, we have used information from the Mexico ENAPROCE survey and show that the index that emerges from the small number of variables collected is highly correlated with the full MOPS index and it captures a large fraction of the cross-firm variance in the quality of management practices.[60]

To investigate also the potential role of policies on technology adoption, the survey asks questions about awareness about existing public programs to support technology upgrading;

[60]Specifically, we use data from Mexico ENAPROCE survey and calculate the correlation between a management quality index with the 4 questions in FAT and the overall index using all questions of MOPS that are in ENAPROCE. The correlations are 0.74 for 2015 survey and 0.73 for 2018; which suggests that with less questions we are still able to capture most of the variation in management quality.

and whether the firm is a beneficiary of such programs, and if so, what type of support the firm received.

While the approach of the survey is as much as possible to ask factual questions, it is also important to understand the perceptions that entrepreneurs and managers have about what are the main barriers and drivers of the decisions to adopt new technologies. To this end, the questionnaire asks the respondent to select the most important obstacle and driver for adoption from a closed list of options. As barriers we include: lack of information and technical skills, uncertainty about demand, cost, lack of finance, government regulations or lack of infrastructure. As drivers, we include competition, adoption by other firms, production of new products, accessing new markets, cost reductions or adjusting to regulations. The survey also asks managers to benchmark their business technology sophistication level in relation to other firms in the country, and also vis-a-vis more advanced firms internationally. This helps to understand the role of beliefs of the main managers in technology adoption decisions and potential behavioural biases and overconfidence.

A.1.5 Balance Sheet

In addition to the information on the technologies used by firms, the survey also collects balance sheet information, information on the business owners, employees, and on potential drivers of and barriers to technology adoption.

Balance sheet. The survey asks the establishment about its total sales, material inputs, replacement value of capital stock, energy consumption, wages and employment. This allows to construct measures of nominal value added per worker, and capital per worker.

Employment. Beyond the number of employees, the survey asks questions that provide information on the education of the workers (share of workers with primary, secondary and tertiary education), and about the occupation composition of the labor force (share of Managers, Professionals, and Technicians; Clerical support workers and sales workers; Production workers and Service workers).

A.2 Sampling frame

The sampling frames were based on the most comprehensive and latest establishment census available from national statistical agencies or administrative business register. Table A.1 provides the main data sources used in the sample frame for each country.

Table A.1: Sampling frame by country

Country	Source	Sampling frame	Year
Bangladesh	Bangladesh Bureau of Statistics.	Est. census, 2013	2019
Brazil	Ministry of Labor	Employer census, RAIS, 2018	2019
Burkina Faso	Business Registry	Business Registry	2021
Cambodia	Tax Registry	Tax Registry	2022
Chile	Business Registry	Census on Establishments	2022
Croatia	Financial Agency (FINA)	FINA Data	2023
Ethiopia	Ministry of Trade and Industry (MoTI)	Business Registry	2022
Georgia	National Statistics Office of Georgia	Est. census, 2021	2022
Ghana	Ghana Statistical Service	Est. census, 2013–18	2021
India	Central Statistics Office of India	Est. census, 2013–17	2020/23*
Kenya	Kenya National Bureau of Statistics	Est. census, 2017	2020
Korea, Rep.	Statistics Korea	Est. census, 2018	2021
Poland	Statistics Poland	Est. census, 2020	2021
Senegal	National Agency for Statistics (ANSD)	Est. census, 2016	2019
Vietnam	General Statistics Office of Vietnam	Est. census, 2018	2019

Note : * The states of Tamil Nadu and Uttar Pradesh were surveyed in Wave 1 in 2020. The states of Gujarat and Maharashtra were surveyed in 2023.

The universe of study includes establishment with 5 or more employees in agriculture, manufacturing and services. The sector classification is based on the International Standard Industrial Classification of All Economic Activities (ISIC), Rev. 4. More specifically, our sample includes firms from the following ISIC rev 4 sectors: Agriculture (ISIC 01, from Group A); All manufacturing sectors (Group C); Construction (Group F), Wholesale and retail trade (Group G), Transportation and storage (Group G), Accommodation and food service activities (Group I), Information and communication (Group J), Financial and insurance activities (Group K), Financial services (ISIC, 64), Travel agency (ISIC 79, from group N), Health services (ISIC 86, from group Q), and Repair services (ISIC 95, from Group S).

Table A.2: Total number of firms in the universe covered by the survey

Country	**Total**	**Sector**			**Size**		
		Agri.	**Manu.**	**Serv.**	**Small**	**Medium**	**Large**
Bangladesh	15,358		15,358		4,164	3,425	7,769
Brazil*	23,364	392	4,758	18,214	12,771	8,955	1,638
Burkina Faso	57,328	4,808	7,493	45,027	40,189	13,284	3,855
Cambodia	8,172		1,890	6,282	5,842	1,287	1,043
Chile	104,854	7,419	11,943	85,492	65,425	30,071	9,358
Croatia	22,350	524	5,387	16,439	17,038	4,381	931
Ethiopia	144,583	3,670	6,553	134,360	105,038	36,798	2,745
Georgia	14,839	313	2,194	12,332	10,815	3,259	765
Ghana	42,165	880	10,284	31,001	30,133	10,070	1,962
India**	616,833	71,464	233,684	311,685	624,452	70,928	13,514
Kenya	74,255	3,680	5,407	65,168	50,584	16,676	6,995
Korea, Rep.	545,515	1520	167,466	376,529	450,264	82,403	12,848
Poland	244,983	3,021	52,340	189,622	198,107	37,799	9,077
Senegal	9,583	1,051	4,069	4,463	7,805	1,414	364
Vietnam	179,713	1,080	45,805	132,828	135,046	33,107	11,560
Total	**2,103,895**	**117,070**	**567,829**	**1,432,427**	**1,756,723**	**350,188**	**75,612**

Note : * Brazil refers to state of Ceará; ** States of Tamil Nadu, Uttar Pradesh, Gujarat, and Maharashtra in India. The survey does not cover agriculture or services in Bangladesh, nor agriculture in Cambodia. In India, only the states of Gujarat and Maharashtra have agriculture included in the survey. Table 1 provides the distribution of the number of firms sampled in each country, by sector and firm size group.

We exclude micro-firms with fewer than 5 employees. Micro firms, particularly in developing countries, are more likely to be informal (Ulyssea, 2018), making them less likely to be captured in the sampling frame; and this would require further adjustment in the survey instrument and sampling design.[61] This size threshold is aligned with other firm-level standardized surveys with comparability across countries. The World Bank Enterprise Survey (WBES) also uses a threshold of 5 employees. The World Management Survey (WMS) uses a threshold of 50 employees.

We stratify the universe of establishments by firm size, sector of activity, and geographic regions. Our sample is representative across these dimensions. In the firm size stratification, we have three strata: small firms (5-19 employees), medium firms (20-99 employees), and large firms (100 or more employees). Regarding sector, for all countries, we stratified at least for agriculture (ISIC 01), food processing (ISIC 10), Wearing apparel (ISIC 14),

[61] In addition, establishments below this threshold often lack the organizational structure to respond to some of the questions.

Retail and Wholesale (ISIC 45, 46 and 47), other manufacturing (Group C, excluding food processing and apparel), and other Services (including all other firms, excluding retail). We use this sector structure of the data for most of the analysis in this paper. Additional sector stratification that were country specific included: motor vehicles (ISIC 29); Leather (ISIC 15), Pharmaceutical (ISIC 21), and Motor vehicles (ISIC 29); and Land transport (ISIC 49), Finance (ISIC 64), and Health (ISIC 86).[62] In the geographic stratification, we use sub-national regions.

To calculate the optimal distribution of the sample, we followed a similar methodology as described by the World Bank (2009). The sample size for each country was aligned with the degree of stratification of the sample.

The data used in this paper corresponds to the first and second phase of the survey implementation. The surveys were administered between June 2019 and the end of 2021 by the World Bank in partnership with public or private local agencies across ten countries: Bangladesh, Brazil (the state of Ceará), Senegal, and Vietnam in the first phase until January 2020. In the second phase, conducted during the COVID-19 pandemic, after January 2020, included Burkina Faso, India (the states of Tamil Nadu and Uttar Pradesh), Ghana, Kenya, Poland, and the Republic of Korea. The mode of data collection was face-to-face before the pandemic and mostly on the telephone during the pandemic.

Table A.3: Year and mode of data collection

Country	Year	Mode
Bangladesh	2019	Face-to-face
Brazil	2019	Face-to-face
Burkina Faso	2021	Telephone
Cambodia	2022	Telephone
Chile	2022	Telephone
Croatia	2023	Online
Ethiopia	2022	Face-to-face
Georgia	2022	Online & Telephone
Ghana	2021	Telephone
India	2020/23*	Face-to-face
Kenya	2020	Telephone
Korea, Rep.	2021	Telephone
Poland	2021	Telephone
Senegal	2019	Face-to-face
Vietnam	2019	Face-to-face

[62] These specific stratifications were taken into consideration when determining sampling weights.

A.3 Survey Weights

We construct the sampling weights of establishments in two steps. First, we compute design weights as reciprocals of inclusion probabilities. Then, to mitigate the risk of non-response bias, we adjust the design weights for non-response.

We adopt a stratified one stage element sampling design and randomly select establishments with equal probabilities within strata. Therefore, the inclusion probability of establishment k, within stratum isr (identified by industry i, size s, and region r), is:

$$\pi_{isrk} = \frac{n_{isr}}{N_{isr}} \tag{A.1}$$

where n_{isr} is the number of establishments targeted by the survey for stratum isr, and N_{isr} is the number of establishments in the sampling frame for the same stratum. Accordingly, the design weights of establishments are:

$$d_{isrk} = \frac{1}{\pi_{isrk}} = \frac{N_{isr}}{n_{isr}} \tag{A.2}$$

To adjust the design weights in Equation A.2 for non-response we follow a simple Response Homogeneity Groups (RHG) approach (Särndal, Swensson and Wretman, 1992), with the groups determined by the strata. In other words, we assume that establishment response probabilities are the same within each stratum, but differ across different strata. Under the RHG approach assumptions, response probabilities can be estimated using the observed response rates within each group, and bias protection is obtained by dividing design weights by group-level response rates.

Denoting with the estimated response probability in stratum isr, and with m_{isr} the number of respondent establishments in the stratum (so that $m_{isr} n_{isr}$), the non-response adjusted weights can thus be written as follows:

$$w_{isrk} = \frac{d_{isrk}}{\hat{\theta}_{isr}} = \frac{d_{isrk}}{m_{isr}/n_{isr}} = \frac{N_{isr}/n_{isr}}{m_{isr}/n_{isr}} = \frac{N_{isr}}{m_{isr}} \tag{A.3}$$

Note that the adjusted weights in Equation A.3 are such that the distribution of our respondent sample across strata exactly matches the distribution of establishments in the sampling frame:

$$\sum_{k \in R_{isr}} w_{isrk} = N_{isr} \tag{A.4}$$

where R_{isr} denotes the respondent sample for stratum isr.

Because of the different number of establishments in each country, when computing global statistics, we re-scale weights so that all countries are equally weighted.

A.4 Measures to minimize bias and measurement error during survey design and implementation

During the design of the survey questionnaire a number of good practices were considered in order to minimize different types of potential biases. The literature on survey design has identified three types of potential bias and measurement errors. These depend on whether they originate from the non-response, the enumerator or the respondent (Collins, 2003). In this section we describe all the steps taken in the design and implementation of the FAT survey to minimize these errors.

Non-response bias. A critical potential bias is associated with non-response in particular questions or non-participation in the survey (Gary, 2007). When this non-response follows a pattern that can be linked to factors correlated to the measured object, this non-response is associated with biases. For example, if more technology sophisticated firms refuse to participate because of fear to reveal commercial information, this would result in significant downward bias in estimating the level of technology sophistication. To minimize this risk, we try to maximize participation in the survey and follow three steps. First, we partner with national statistical offices and industry associations to use the most comprehensive and updated sampling frame available, as well as their experience on data collection, which are supported by endorsement letters from local institutions.[63] Having up to date contact details significantly improves response and minimized contact fatigue. Second, we follow a standard protocol in which each firm is contacted several times to schedule an interview. We split the sample in different batches, following the order of randomization within stratum, and provide contact information of subsequent batches only after interviewers have shown evidence that they have exhausted the number of attempts to complete the initial list. Third, we monitor the implementation, validation of skip conditions and outliers (e.g. balance sheet information) in real time using standard survey software, and request that any missing information are completed through a follow up call, checked by supervisors. This minimizes risks that enumerators skip the order of their randomly assigned list of firms.

Enumerator bias and error counts. Minimizing cognitive biases in respondents in face to face and phone interviews starts with making sure that enumerators are able to implement the survey in a clear and consistent manner. To this end, the survey, training, and data collection processes are largely designed to minimize enumerator biases and data collection

[63]These procedures are in line with suggestions of good practice for implementation by (Bloom et al., 2016).

errors. First, to reduce the likelihood of coding errors, we use closed-ended questions, which make coding the answers a mechanical task, eliminating the reliance on the enumerator's interpretation of the answer and subjective judgement to code them, as it is the case with open-ended questions (Bloom et al., 2016). Second, to make sure that implementation is consistent across enumerators within and between country surveys, we implement the same standardized training in each country with enumerators, supervisors, and managers leading the data implementation. The training is led by team members directly involved in the elaboration of the questionnaire and implemented in local languages - English, French, Portuguese and Vietnamese,[64] and they include vignettes to ensure that enumerators understand the specific technologies they are asking about. The two to three days training consists of one general presentation about the project, covering the main motivation, relevance, coverage, and protocols that should be used to approach the interviewees and the review of the full questionnaire (question by question). The training material includes pictures of each technology mentioned in the survey both in general and sector-specific business functions, which are shared with enumerators. After going over the full questionnaire and clarifying any questions that emerge, the participants of the training conduct a mock interview using CAPI, under the supervision of our team.

Third, to guarantee that translations use words that are understood by firms managers, in each country we conduct a pre-test pilot of the questionnaire with firms out of the sample. A pilot of the questionnaire is implemented in each country with firms out of the sample. This allows to fine-tune questions to the local language, finalize the translation and select the most relevant examples in each question. After the pilot, our teams have the opportunity to discuss with the managers implementing the questionnaires and clarify any potential question over the implementation process.

Fourth, to attain greater quality control during the data collection process, enumerators record the answers via *Computer-Assisted Personal Interviews* (CAPI) and *Computer-Assisted telephone Interviews* (CATI) software.[65] Using CAPI/CATI has clear advantages. First, it allows the use of logical conditions and skips which prevent data inputting errors and omitting questions, and also reduces the potential for abnormal values or non-response to specific questions. Second, it reduces substantially the time of implementation of the survey, increasing the quality of responses and minimizing survey fatigue. Supervisors are assigned to review all interviews, identifying missing values and abnormal responses. In addition, the CAPI/CATI system can identify when enumerators complete the survey too fast and

[64] In the case of Vietnamese, we used translation services support.

[65] Randomized survey experiments with household survey has demonstrated that a large number of errors observed in *Pen-and-Paper Personal Interview* (PAPI) data can be avoided in CAPI (Caeyers, Chalmers and De Weerdt, 2012).

other abnormal issues that can raise concerns about the quality of the interview. Finally, CAPI/CATI also allows for the core team to regularly monitor the data collection process and use standard algorithms to analyze the consistency of the data at different stages of data collection and by watches, thus providing continuous feedback and quality control.

Respondent bias. Perhaps the most important type of bias relates to cognitive biases from respondents. These biases can be large in surveys with open ended questions or where concepts can be largely subjective. Specifically, two broad groups of factors can trigger response errors: *cognitive*, which affect the comprehension of the questions, and *framing*, which may cause biased answers due to the perceived socially (un)desirability of the answers (Bertrand and Mullainathan, 2001). We take several steps to minimize this respondent bias. First, surveys need to be responded by the appropriate person in the firm that has all the information needed to respond. During the implementation of the screening process we ensure that the interview is arranged with the appropriate person or persons (Bloom et al., 2016). Senior managers (and in larger firms other managers such as plant managers) are asked to respond to the sections that cover the technologies used, and HR managers are asked to respond the questions on employment. Second, when possible use face-to-face interviews, which lead to higher response rates and lower respondent bias and measurement errors than web-based interviews.Only during the pandemic and due to existing mobility restrictions, we implemented surveys on the phone. Third, as discussed above, the use of a closed-ended design in the questionnaire reduces measurement error in the answers as the respondent is questioned about specific technologies (one at a time), and only when the presence of each of the possible technologies is established, the question about the most widely used technology is triggered. While this increases the length of the interview, it also increases the reliability of the data collected. Fourth, also as discussed above, we pre-pilot the questionnaire in each country to ensure that questions are clear in their wording in the specific geographical and cultural contexts, simple, and objective, so that the response does not require any subjective judgement (Bertrand and Mullainathan, 2001). Fifth, and more importantly, to avoid *social desirability bias*, by which respondents may overstate the use of more sophisticated technologies, the survey avoids the words "technology" and "sophistication" and employs more neutral terms such as "methods" and "processes". In addition, the survey is administered so that the respondent does not know all the possible technologies in a business functions before asserting whether a technology is used in the firm.[66] This reduces the risk that managers are framed to bias responses to the more advanced (socially desirable) technology, since they don't know what they will be asked in advance. Finally, when possible, enumerators are instructed to visually verify the information provided during the interviews. For example,

[66] It also allows for "don't know" options.

in the case of use of a sophisticated production technology that can be visually identified in the shop floor.

A.5 Ex-post checks and validation exercises

In addition to using best practices in survey design and implementation, it is important to perform validation checks once the data is collected. This allows us to measure the effectiveness of all these efforts to minimize bias and measurement error. In what follows, we describe some of the validation tests performed.

Minimizing potential non-response bias Our survey implementation was designed to minimize non-response through the use of well-prepared agencies and institutions to administer the survey and the presentation of adequate supporting letters to encourage participation. Table A.4 shows response rates by country, firm size group and sector. Response rates vary between 15% in Croatia and 86% in Georgia.

Table A.4: Response rates (by country)

Country	Response rate
Bangladesh	30%
Brazil	39%
Burkina Faso	45%
Cambodia	16%
Chile	40%
Croatia	15%
Ethiopia	42%
Georgia	86%
Ghana	49%
India	49%
Kenya	77%
Korea*	24%
Poland	47%
Senegal	57%
Vietnam	80%
Average across countries	**46%**

These are unweighted response rates calculated as the ratio between firms that responded the survey and the total number of firms in the sample which we attempted to conduct the interview. The high response rate for Vietnam is associated with the fact that the survey was implemented by the national statistical office. In most cases, these response rates are high relative to typical response rates in firm-level surveys, which for the U.S. are around 5 to 10 percent, and are consistent with response rates observed for WMS and MOPS (Bloom et al., 2016).[67]

To minimize potential non-response bias, we adjusted the sampling weights for unit non-response. The non-response adjustment was calculated at the strata level, so that the weighted distribution of our respondent sample across strata (sector, size, region) exactly matches the distribution of establishments in the sampling frame.

More importantly, to check the reliability of the instrument we implemented a series of ex-post tests in the first phase of the survey, focusing on countries we implemented the survey first. First, we study whether, in the sample of contacted firms, there are significant differences between those that responded and those that declined participating or could not be reached. The only information available in all firms we attempted to contact in the three sampling frames is the number of employees. Table A.5 tests whether there are differences in employment between the respondent and non-respondent groups, controlling for characteristics used for stratification. We find no significant differences in firm size between respondents and non-respondents in any of the three countries.

Table A.5: Comparison of establishment size between respondents vs non-respondents

VARIABLES	Brazil	Vietnam	Senegal
Respondents (FAT)	2.52	52.34	-4.92
	(22.19)	(80.27)	(6.63)
Observations	1,754	1,500	3,075
R-squared	0.129	0.172	0.237
Controls:			
Sector FE	Y	Y	Y
Size-group FE	Y	Y	Y
Region FE	Y	Y	Y

Note : *** p<0.01, ** p<0.05, * p<0.1. Data are from the list of establishments contacted by the enumerators. For each country, the level of employment was regressed on a dummy for respondent while controlling for stratification such as sectors, size groups (small, medium, and large), and regions. Estimates for Vietnam are based on the original list of 1500 firms, with 1346 respondents and 154 non-respondents. Robust standard errors in parenthesis.

Second, under the premise that any systematic relationship between firm characteristics and participation is continuous in their reluctance to participate in the survey, we can learn about sample differences between respondents and non-respondents by comparing firms across different percentiles of the distribution of the number of attempts it took for them

[67]The average response rate for the WMS is around 40 percent.The response rate for MOPS, implemented by the United States Census Bureau, was around 80 percent.

to respond the survey.[68] For Senegal, we explore whether after controlling for observable characteristics, there are significant differences in average technology sophistication in GBFs between firms that required a larger number of attempts to be contacted (top quartile) and those that did not. Table A.6 shows that there are no statistically significant differences in technology sophistication between the two groups.

Table A.6: Comparison of technology sophistication between high and low number of attempts

VARIABLES	Senegal	Senegal
Top quartile of attempts (4 or more)	-0.021	-0.027
	(0.020)	(0.019)
Observations	1,753	1,666
R^2	0.377	0.437
Controls:		
Sector FE	Y	Y
Size-group FE	Y	Y
Region FE	Y	Y
Age		Y
Exporter		Y
Foreign owned		Y

Note: *** p<0.01, ** p<0.05, * p<0.1. Data are from the Senegal FAT survey with information on the number of attempts to complete interview at the firm level. Technology sophistication is regressed on a dummy for the top quartile of the number of attempts (4 or more) with controls for the stratification (sectors, size groups, and regions) and/or firm characteristics (age groups, exporter, and foreign owned). Robust standard errors in parenthesis.

Third, we compare firms that were in the first sample list provided to enumerators and those in subsequent lists. Table A.7 show that there are no statistically significant differences between the two groups.

In each of these exercises, we find no statistical difference in the number of employees, technology sophistication, wages, and share of workers by skill and education between firms in the group that proxies for the response sample and the group of firms that proxies for the non-response sample.[69]

Minimizing enumerator bias. To minimize the potential for enumerators to introduce biases when administering the survey, we conduct in each country the same standardized training and piloting prior to going to the field. We also conduct ex-post tests to identify

[68] Behaghel et al. (2015) infer the reluctance to participate in the survey from the number of attempts that it take for a firm to accept the request.

[69] See Table A.5 to A.11 in Appendix A.

Table A.7: Comparison of technology sophistication between original and replacement sample

VARIABLES	Brazil	Brazil	Vietnam	Vietnam	Senegal	Senegal
Original sample	-0.014	-0.037	0.030	0.043	0.021	0.028
	(0.048)	(0.047)	(0.050)	(0.048)	(0.018)	(0.018)
Observations	638	637	1,484	1,484	1,753	1,666
R-squared	0.299	0.335	0.262	0.320	0.377	0.437
Controls:						
Sector	Y	Y	Y	Y	Y	Y
Size group	Y	Y	Y	Y	Y	Y
Region	Y	Y	Y	Y	Y	Y
Age		Y		Y		Y
Exporter		Y		Y		Y
Foreign owned		Y		Y		Y

Note: *** p<0.01, ** p<0.05, * p<0.1. Data are from the Brazil, Vietnam, and Senegal FAT surveys. For each country, technology sophistication ($MOST_j$) is regressed on a dummy for the original sampling list with controls for the stratification (sectors, size groups, and regions) and/or firm characteristics (age groups, exporter, and foreign owned). Robust standard errors in parenthesis.

abnormal interviews or outliers by running regressions of firm-level sophistication on enumerator dummies and firm controls as discussed in the text. Table A.8 shows that enumerator dummies are not significant for Brazil, Ghana, India, and Korea. For Bangladesh, Kenya, Senegal, and Vietnam, less than 12% of enumerator dummies are statistically significant. Table A.9 compares the average technology sophistication ($MOST_j$) for GBF, excluding the firms with abnormal enumerators and in the entire sample. We find no economic or statistical difference between mean sophistication in these countries.

Minimizing respondent bias. A critical factor to minimize respondent bias is to identify the right respondent. The protocol for the implementation of the survey required that the survey should be ideally answered by the top manager. About 47% of the survey was answered by the owner or CEOs, while the other responses included factory managers, other managers, administrative staff, and accountants. Almost 80% of the interviews were conducted through one visit in person interview with the main respondent. In circumstances in which the main respondent did not have all the information about a general topic of the questionnaire, especially in modules B and C, they were requested to consult with other colleagues.

To assess the relevance of response bias, we conduct a parallel pilot in Kenya where we re-interview 100 randomly selected firms with a short version of the questionnaire. For

Table A.8: Analysis of enumerator bias

VARIABLES	Brazil	Vietnam	Senegal	Bangladesh
Share of Significantly Different Interviewers	0	0.09	0.08	0.11
Number of Significantly Different Interviewers	0	13	2	4
Number of Interviewers	8	145	25	37
	Ghana	India	Korea	Kenya
Share of Significantly Different Interviewers	0	0	0	0.006
Number of Significantly Different Interviewers	0	0	0	9
Number of Interviewers	44	18	9	450

Note: Data from the Firm-level Adoption of Technology (FAT) surveys in Brazil, Vietnam, Senegal, Bangladesh, Ghana, India, Korea and Kenya. Significantly different interviewers are identified from the regressions of employment on interviewer dummies with controlling for stratification information (e.g., sector, size, and region). For each country, the share of significantly different interviewers is computed by dividing the number of interviews conducted by significantly different interviewers by the total number of interviews.

those firms, we randomly select three business functions and ask about the presence of the relevant technologies.[70] Both the original and back-end interviews in the pilot are conducted by phone by different interviewers.

Despite using phone interviews which are subject to greater measurement error than face-to-face interviews, comparison of answers from the pilot reveals that 73% of the answers were the same across the two interviews.[71] We estimate a probit model to assess the likelihood of consistent answers between the original and the back-check interviews, controlling for firm-level fixed-effect. Reporting the use of a technology in the back-check interview is associated with 80.6% of likelihood of reporting the use of the same technology in the original interview. Conversely, reporting that a technology is not used in the back-check interview, is associated with a 29.3% likelihood of being reported in the original survey.

Additional validation exercise with employer-employee census (RAIS) in Brazil

Some final ex-post checks were conducted with the Brazil data and takes advantage of the fact that we have access to the RAIS administrative data, which is a matched employer-employee dataset that covers the universe of firms in the sampling frame. This allows us to compare variables in RAIS with variables we collected in FAT for the same firms.

[70] The pilot coincided with the beginning of the data collection for phase two which includes new countries, and Kenya is one of them. Despite the fact that Kenya is not in the sample, the pilot is informative about the significance of response bias. The re-interviews produce 1,661 answers (106 interviews times 3 business functions times an average of 5.2 technologies per function).

[71] The consistency ranges from 65% in business administration to 77% in sales across business functions, and from 85% among the most basic technologies to around 61% in intermediate, and 77% at the most advanced technologies across functions.

Table A.9: Difference in technology sophistication in general business functions with and without outlying enumerators

	All Sample	Sample Without Different Enumerators	Difference
Vietnam			
Mean	1.934	1.947	-0.013
SE	(0.012)	(0.012)	(0.017)
Observations	1,499	1,341	
Senegal			
Mean	1.406	1.404	0.002
SE	(0.011)	(0.011)	(0.016)
Observations	1,786	1,784	
Bangladesh			
Mean	1.482	1.458	0.024
SE	(0.015)	(0.015)	(0.021)
Observations	903	798	
Kenya			
Mean	1.938	1.936	.002
SE	(0.020)	(0.020)	(0.029)
Observations	1305	1296	

Note: *** p<0.01, ** p<0.05, * p<0.1. Data from the Firm-level Adoption of Technology (FAT) surveys in Vietnam, Senegal, Bangladesh and Kenya. Brazil, Ghana, India and Korea are excluded because they do not include significantly different interviewers. The average of technology sophistication in general business functions ($MOST_j$) is compared between all sample and sample excluding significantly different enumerators. Standard errors in parenthesis.

First, we analyze the correlation between sales per worker and our technology measures (GBF) and (SBF) from FAT and average wages from RAIS. Table A.10 reports the point estimates of regressing firm-level FAT variables on the log or average wages per worker from RAIS and a set of firm-level controls. The FAT variables are log of sales per worker (column 1), and average technology sophistication (GBF, column 2, and SSBF, column 3). In all three cases we find strong positive associations between the FAT and the RAIS variables.

Second, we compare the differences between labor-related indicators from a matched employer-employee administrative data for firms in FAT versus the universe of firms. To perform this comparisons we obtained the weighted average for firms in FAT, using the weights we constructed as described in section A3 and compare it with the average for all firms in RAIS that are part of our universe for the State of Ceará, in Brazil [72]. We then

[72]The variables are number of workers, average wages, share of workers with college degree, share of low

Table A.10: Relationship between FAT survey variables and log of wages from administrative data for Brazil

Variable	(1) log(sales per worker)	(2) GBF	(3) SSBF
ln(Wage) RAIS	0.882***	0.400***	0.299***
	(0.157)	(0.111)	(0.101)
Observations	592	675	674
R-squared	0.346	0.364	0.800
Controls:			
Sector FE	Y	Y	Y
Region FE	Y	Y	Y
Size-group FE	Y	Y	Y
Age	Y	Y	Y
Exporter	Y	Y	Y
Foreign owned	Y	Y	Y

Note: *** p<0.01, ** p<0.05, * p<0.1. Average wage information for each establishment is obtained from the 2017 *Relação Anual de Informações Sociais* (RAIS) merged with the Firm-level Adoption of Technology (FAT) data used in this exercise, including sales per worker, the technology adoption index ($MOST_j$) for GBF and SSBF, and firm characteristics used as controls. Regressions estimated using establishment-level sampling weights. Robust standard errors in parenthesis.

perform a t-test to compare the differences. Table A.11 shows that the differences are not statistically significant.

Overall, these ex post checks appear to validate the quality of the data collected.

skilled, and share of high-skilled workers, where high- and low-skilled workers are defined as in Autor and Dorn (2013).

Table A.11: Comparison between FAT sample and RAIS data (universe)

	Number of employees	Average wage	Share college	Share low-skill	Share high high-skill
FAT Average (weighted)	28.55	1,311.89	0.05	0.16	0.42
RAIS Average (universe)	23.85	1,349.29	0.05	0.17	0.39
Estimate (RAIS - FAT)	-4.70	37.40	0.00	0.00	-0.03
Standard Error	(3.08)	(29.77)	(0.01)	(0.01)	(0.02)
T-Statistic	-1.52	1.26	0.55	0.20	-1.64

Note: *** p<0.01, ** p<0.05, * p<0.1. Data from the 2017 *Relação Anual de Informações Sociais* (RAIS) and the Firm-level Adoption Technology (FAT) survey in Brazil. The estimates from RAIS data are unweighted, and those from FAT surveys are weighted by the sampling weights. Robust standard errors in parenthesis.

B Construction of Measures

This section explains in detail, the construction of the various technology sophistication and task-based specialization measures discussed in the paper. It also details the all the exceptions where the construction of the variables is deviated from the norm.

B.1 Technology Measures

The technology measures at the BF-level that are discussed in the text are as follows - N_f, $ANUM_{f,j}$, $NUM_{f,j}$, $MAX_{f,j}$, and $MOST_{f,j}$. The means of all of these BF-level measures are reported in Table C.18. To understand the construction of these variables, it's imperative to understand the structure of the corresponding questions asked. The figure below elucidates an example question asked for a particular business-function (BF) in the "Livestock" Sector. The presence of each of the technologies is calculated as a binary variable taking values 0 and 1[73].

Question: Does this establishment use any of the following methods for breeding or genetic procedures?	Answer	
Breed substitution	1 = Yes 0 = No	. = Don't know
Inbreeding or Crossbreeding	1 = Yes 0 = No	. = Don't know
Artificial insemination (AI)	1 = Yes 0 = No	. = Don't know
Selective breeding based on molecular genetics (marker-assisted selection, PCR, DNA sequencing, transgenesis, cloning, genomics)	1 = Yes 0 = No	. = Don't know
Other	1 = Yes 0 = No	. = Don't know

Question: What is the breeding or genetic method used more often in this farm?	Answer 1 = Breed substitution 2 = Inbreeding or Crossbreeding 3 = Artificial insemination (AI) 4 = Selective breeding based on molecular genetics (marker-assisted selection, PCR, DNA sequencing, transgenesis, cloning, genomics) . = Other . = Don't know

Figure B.14: Example question for the presence of technologies and most-used technology

If a respondent answers that they use more than one technology for the BF, then they

[73]If a respondent answers "Don't know", that is coded as missing. We also do not take into account for technologies outside the grid (Other).

are asked about the most-used technology in that BF[74].

After the construction of these technology-level binary variables and the variable for most-used technology (which is at Establishment-BF level), we move on to constructing the N_f, $ANUM_{f,j}$, $NUM_{f,j}$, $MAX_{f,j}$, and $MOST_{f,j}$, variables. The details and formulae used to construct each of these variables is listed below -

N_f : This simply denotes the number of technologies in the grid that exist for business function f. This does not depend on the answers provided by the survey respondents, and takes the same value for a particular BF, across all establishments[75].

$ANUM_{f,j}$: $ANUM_{f,j}$ denotes the "Absolute" Number of Technologies used by an establishment in a particular business function. To construct this, we just take into account the number of technologies in a BF, that the establishment said "Yes" to using. Note that the $ANUM_{f,j}$ would *always* be lesser than N_f.

$NUM_{f,j}$: This denotes the "Relative" Number of Technologies. To calculate this the following formula is used: $NUM_{f,j} = \frac{ANUM_{f,j}-1}{N_f-1} * 4 + 1$. This describes a scaled measure for the number of technologies used in an establishment. Both, the numerator and denominator are subtracted by 1 in order to assign a 0 (which then gets scaled up to 1), for an establishment that uses just 1 technology in a BF. The scaling is done as such that the range of the values is from $[1, 5]$ for all business functions across establishments, hence allowing for comparability.

$MAX_{f,j}$: $MAX_{f,j}$ denotes the relative ranking of the most-sophisticated technology used by the establishment in the BF. To calculate this the following formula is used: $MAX_{f,j} = \frac{r_{f,j}^{MAX}-1}{R_f-1} * 4 + 1$. $r_{f,j}^{MAX}$ is the rank of the most sophisticated technology used in a BF. R_f is the rank of the most sophisticated technology **possible** in the BF f. For majority of the business functions, R_f would simply coincide with N_f[76]. However, there are particular exceptions in BFs where the number of technologies in the grid for the BF is not equal to the rank of possibly most sophistication technology. These are discussed in **??**. The ranks are again adjusted by a constant scaling factor and resultant $MAX_{f,j}$ values are in the range of $[1, 5]$.

[74] Again here, if the answer is "Don't know", or "Other", we assign it a missing value.

[75] For instance in the example in Figure B.14, the value that N_f takes here is 4 for all establishments. Again, this does not include the "Other" technology, as it is not specified in the grid.

[76] For instance, in the example question if a respondent answered "Yes" for "Breed substitution", "Inbreeding", and "Artificial Insemination", and "No" for all other choices, $r_{f,j}^{MAX}$ would take value "3" for that establishment for this particular BF. Here $R_f = N_f = 4$. Hence, the calculated value of $MAX_{f,j}$ would be "3.67".

$MOST_{f,j}$: $MOST_{f,j}$ denotes the relative ranking of the most-used technology in a particular BF by the establishment. To calculate this the following formula is used: $MOST_{f,j} = \frac{r_{f,j}^{MOST}-1}{R_f-1} * 4 + 1$. $r_{f,j}^{MOST}$ is the rank of the most used technology used in a BF. Definition of R_f stays the same as above[77]. The ranks are again adjusted by a constant scaling factor and resultant $MOST_{f,j}$ values are in the range of $[1, 5]$

Now after constructing these measures at the establishment-BF level, these are averaged out at the establishment-level to get the measures NUM_j, MAX_j, and $MOST_j$, calculated as follows -

$$S_j = \sum_{f=1}^{N_j} \frac{S_{f,j}}{N_j} \tag{B.5}$$

where $S \in \{NUM, MAX, MOST\}$, and N_j is the number of BFs conducted in establishment j.

B.1.1 Exceptions

As mentioned above, there are some exceptions to the calculation of $MAX_{f,j}$ and $MOST_{f,j}$ statistics for particular business functions. These exceptions are limited to two sectors - Automotives (Motor Vehicles) and Health.

Automotives - As specified in Figure A.7, in BFs "Body Pressing", "Painting", and "Plastic Injection Molding", although the number of technologies is more than 2, there are sub-BFs, where the ranking of technologies is only 2 (basic and advanced). For instance, in "Body Pressing", there are three distinct sub-BFs - "pressing of skin panels", "structural components", and "welding of main body" (refer to Figure A.7). As such, if the ranking of the technologies were to take the value of the technology as per the questionnaire, that would be erroneous. Another example is provided in Figure B.15 for the question related to Body Pressing. Consider for example, an establishment answering "Yes" to 1st, 2nd, 3rd, and 6th technologies, and for intensive margin answering "3".

If we were to calculate scaled-up $MAX_{f,j}$ and $MOST_{f,j}$ values for this establishment in body pressing, as we do for all other BFs, it would be equal to "5" and "2.6" respectively (if we take $R_f = N_f = 6$). However, this is incorrect, as the 1st, 2nd, 3rd, and correspondingly, 4th, 5th, and 6th technologies are related to the parallel sub-BFs respectively. Hence, we assign

[77] For example - an establishment uses both "Breed sub.", "Inbreeding" and "Artifical insem.", but answer that they used "Inbreeding" most often. Then the value $r_{f,j}^{MOST}$ takes is 2. Consequently, $MOST_{f,j}$ takes the value "2.33", for this establishment in this BF.

Question: Does this establishment use any of the following body pressing and welding methods?	Answer	
Pressing of skin panels using operators	1 = Yes 0 = No	. = Don't know
Pressing of structural components using operators	1 = Yes 0 = No	. = Don't know
Welding of main body using operators	1 = Yes 0 = No	. = Don't know
Pressing of skin panels automated using robotics	1 = Yes 0 = No	. = Don't know
Pressing of structural components automated using robotics	1 = Yes 0 = No	. = Don't know
Welding of main body automated using robotics	1 = Yes 0 = No	. = Don't know
Other	1 = Yes 0 = No	. = Don't know

Question: What is the pressing and welding method used more often in this facility?	Answer 1 = Pressing of skin panels using operators 2 = Pressing of structural components using operators 3 = Welding of main body using operators 4 = Pressing of skin panels automated using robotics 5 = Pressing of structural components automated using robotics 6 = Welding of main body automated using robotics . = Other . = Don't know

Figure B.15: Question for Extensive and Intensive Margins of Tech. in Body Pressing

rank "1" to the first three technologies, and rank "2" to the last three technologies, and take an arithmetic mean to calculate r_j^{MAX}. So in the example, the most sophisticated technology for the establishment in "pressing of skin" would be rank 1, and both, in "pressing of structural panels" and "welding", would be 2. Effectively r_j^{MAX} would then be an arithmetic mean of $\{1, 2, 2\}$ which is "1.6". Accordingly, value of R_f would be 2, and hence the corrected scaled-up value of MAX and MOST for this estab. in this BF, would be "3.4" and "3" respectively.

Table B.12 shows the adjustment made to the ranks in these BFs before calculation of $MAX_{f,j}$ and $MOST_{f,j}$.

Health - The exceptions in this sector come from two BFs - "Health Equipment" and "Procedures". Firstly, in "Health Equipment" the number of technologies is 11 (see **??**), but there exists no clear ranking. For that reason, to calculate $MAX_{f,j}$ and $MOST_{f,j}$ for this

Table B.12: Adjustment of Ranks in Automotive BFs

Business Function	Sub-Business Functions	Technologies	Rank in Data	Adjusted Rank	R_f	N_f
Body Pressing and Welding	Pressing of Skin Panels	using Operators	1	1	2	6
		using Robots	4	2	2	6
	Pressing of Structural Components	using Operators	2	1	2	6
		using Robots	5	2	2	6
	Welding of Body Parts	using Operators	3	1	2	6
		using Robots	6	2	2	6
Painting	Water-Based	using Operators	1	1	2	4
		using Robots	3	2	2	4
	Solvent-based	using Operators	2	1	2	4
		using Robots	4	2	2	4
Plastic Injection Modelling	Molding of non-visible interior plastic	using Operators	1	1	2	4
		using Robots	3	2	2	4
	Molding of plastic exterior body parts	using Operators	2	1	2	4
		using Robots	4	2	2	4

Notes : The table shows the ranking of technologies in these three BFs of Automotives sector.

BF is difficult. As a result, the value of $MAX_{f,j}$ and $MOST_{f,j}$ is taken to be the same as the value of $NUM_{f,j}$ (which is the relative number of technologies used by an establishment in the BF), for this particular BF.

Coming to "Procedures", both in the data and questionnaire, the questions regarding different procedures are asked individually (see Figure A.12). There are 4 types of procedures - Sepsis Treatment, Childbirth, Trauma, and treatment of Myocardial Infarctions/Strokes. For each of these procedures, there are two corresponding technologies. When the question is asked to the respondent, there are 5 possible options that they could reply, namely - "It is always available", "It is NOT always available", "It is NEVER available", "Don't know", and "Not applicable". For the purpose of classification, firstly the responses for each technology are collapsed into binary variables, taking values 1 for the "always available" option, and 0 for "NOT always" and "NEVER available" options. Just like in other BFs, value of "Don't Know" and "Not Applicable" is considered as missing. After the re-coding, these 4 questions are collapsed into one "Procedures" question that would have 8 technologies. Similar to "Health Equipment", there is no clear ranking to technologies there, hence the value of $MAX_{f,j}$ and $MOST_{f,j}$ is taken to be the same as the value of $NUM_{f,j}$ for this particular BF.

B.2 Task-based Specialization Measures

In this section, we detail the construction of task-based measures of specialization, and measures that define the scope and limits of an establishment. We also talk about the exceptions to the creation of such measures particular to some BFs. Note that only in this section, when we discuss BFs, these are **just SSBFs**[78], as information on in-house, insourcing, and outsourcing isn't asked for GBFs. Hence, the term BFs and SSBFs is used interchangeably, in this section.

The measures discussed in this section are - Irrelevant, Specialization, In-house, Insourcing and Outsourcing.

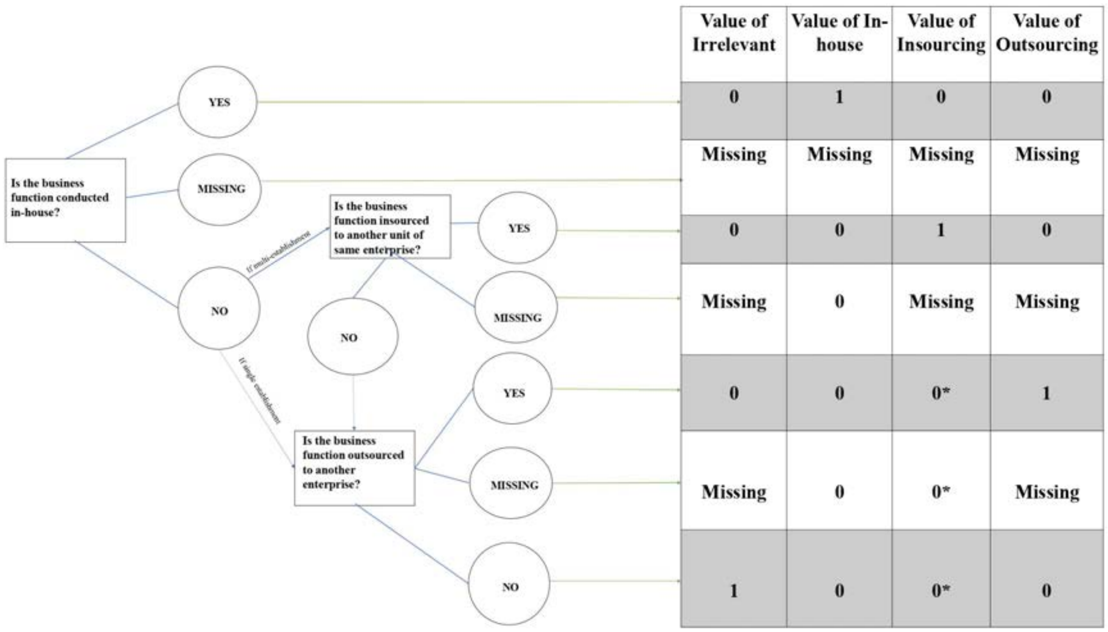

Figure B.16: Construction of Irrelevant, Insourcing and Outsourcing Variables

Notes – The above figure indicates the questions asked for the conduct of business functions, the possible responses, and the coding of Irrelevant, In-house, Insourcing and Outsourcing variables in each of the possible scenarios. The square boxes indicate the questions, followed by the responses in the circles. Blue arrows indicate the process and the flow of questions. Green arrows at the final nodes correspond to the respective values of the different variables under each scenario. * Note that the question for in-sourcing is asked only to multi-establishment firms. As a result, the value of insourcing is "MISSING" for all single-establishment firms and hence, the coding of irrelevant for these firms is decided based only on responses given to the inhouse and outsourcing questions.

Before asking questions on technologies regarding specific BFs, the survey asks whether the BF is conducted in-house, if it is insourced to any other establishment of the same firm (in case of multi-establishment firms), or is outsourced to another firm. The template of such questions, the possible responses, and the subsequent coding of Irrelevant, In-house, Insourcing, and Outsourcing variables is given in the exhibit in Figure B.16. After coding of these four variables at the establishment-BF level, the "Specialization" measure is created.

[78]Here SSBFs does not include Fabrication, for Manufacturing sector establishments

A "Specialized" BF simply indicates a BF that is not performed in-house i.e. it is either irrelevant, insourced or outsourced[79]. These establishment-BF level binary variables are the variables corresponding to $D^{SP}_{f,j}$, $D^{IN}_{f,j}$, and $D^{O}_{f,j}$ that are studied in the paper.

After construction of such variables at the establishment-BF level, each of them is aggregated at an establishment level in two ways - Mean and ANY. Detailed description of each of these variables is given below -

Mean Measures - SP_j refers to the fraction of total BFs in establishment j that are not conducted in-house[80]. IH_j is the fraction of relevant BFs in establishment j that are conducted in-house. IN_j is the fraction of relevant BFs in establishment j that are in-sourced to another unit of same firm/enterprise. O_j is the fraction of relevant BFs in establishment j that are outsourced to other firms[81].

ANY Measures - "ANY" measures are establishment-level binary-variable that take value 1 if the condition is satisfied, and 0 otherwise. ANY^{SP}_j, is a binary-variable that takes value 1 if atleast one of the BFs in an establishment j is not conducted in-house. Similarly, ANY^{IH}_j, ANY^{IN}_j, and ANY^{O}_j take value 1 if establishment j has atleast one relevant BF **and** atleast one of those relevant BFs is conducted in-house, insourced, or outsourced, respectively[82].

After construction of the establishment-level measures these are aggregated using establishment-level sampling weights to report the sector statistics in Table 8 and Table C.26. A BF level correspondence is provided in Table C.28. The exceptions to the above calculations are discussed below -

B.2.1 Exceptions

The exceptions to the construction of above variables comes in some sectors, namely - Agriculture, Pharmaceuticals, and Healthcare.

[79] So it takes value 1 if a BF isn't conducted in-house, and takes value 0, if it is conducted in-house.

[80] Here the denominator is the number of total possible BFs that an establishment could conduct. For instance - if a multi-establishment firm answers in Livestock sector which has 5 SSBFs, answers "Yes" in-house for 1 BFs,"Yes" insourced for 1 BF, "Yes" outsourced for 1 BF, "No" to everything for 1 BF, and doesn't answer anything for 1 BF, then the number of specialized BFs would be 2. SP_j in this case would be $\frac{2}{5}$, because 2 BFs are inhouse, out of total possible 5 BFs an establishment in the Livestock sector could have.

[81] In the example IH_j, IN_j, and O_j would each be $\frac{1}{3}$. Note that IN_j is only calculated for multi-establishment firms. So for single establishment firms, IN_j would just be missing.

[82] In the example, all values : ANY^{SP}_j, ANY^{IH}_j, ANY^{IN}_j and, ANY^{O}_j take value 1.

Agriculture - In Agriculture, the only question asked for Irrigation, is "Whether the product requires Irrigation?", instead of how its conducted. Hence, all those who answered "Yes" to requiring irrigation, are coded as in-house, and "No" as Irrelevant.

Pharmaceuticals - In Pharmaceuticals, the questions on how the BF is conducted are not asked for "Weighing Scale" and "Mixing and Compounding". Hence, the information is missing for these two BFs. Additionally, the questions on the BF "Encapsulation" are only asked to the pharmaceutical establishments that produce "Tablets and capsules", or "Creams and ointments". So for other establishments in Pharma., the information would be missing[83].

Health - In Health establishments, questions on the conduct of business functions are not asked for "Infrastructure and Machines" and "Procedures". Additionally, for "Healthcare Management", the only question asked for the conduct is "Does this establishment have a pharmacy to provide medication for patients in house?". Hence the people who answered "No" are classified in two ways - if they belong to multi-establishment firms, they are coded as "Insourcing", and if they don't belong to multi-estab. firms, they are coded as "Outsourcing". Hence, for this particular BF, Irrelevant always has value 0.

[83] Out of 416 Pharmaceutical Companies, 169 do not produce "Tablets and capsules" nor "Creams and ointments".

C Additional Figures and Tables

C.1 Technological Sophistication

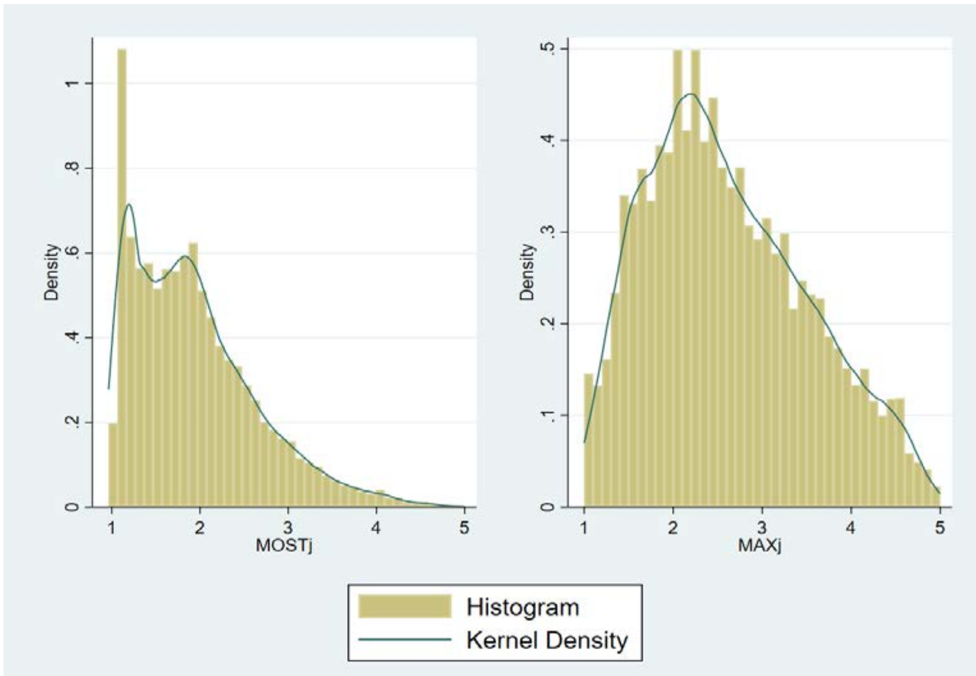

Figure C.17: Distribution of MAX_j and $MOST_j$ at the establishment level

Notes : This figure represents the distribution of sophistication measures at the establishment level - MAX_j and $MOST_j$. The histogram and kernel density curves are calculated using establishment-level sampling weights.

Table C.13: Summary Statistics

	All			Small (5-10)		Medium (11-99)		Large (100+)	
	N	Mean	p50	N	Mean	N	Mean	N	Mean
Total # of employees	18570	33.95	9.00	9909	8.45	5477	38.03	3184	492.04
% of workers with college	17686	30.15	20.00	9520	30.79	5144	27.71	3021	28.12
Management practice (z-score)	20739	0.00	0.27	11029	–0.09	6098	0.25	3610	0.53
Sales per worker	15355	11.45	11.43	8121	11.47	4545	11.39	2689	11.25
	Share			Share		Share		Share	
Multi-establishment	18.4%			12.1%		20.5%		33.9%	
Multinational	17.6%			16.4%		16.4%		23.7%	
Exporter	16.8%			8.1%		18.1%		41.0%	
Age:									
1–5 Years	17.9%			22.2%		15.2%		9.3%	
6–10 Years	21.3%			24.1%		19.6%		15.4%	
11–15 Years	17.9%			18.4%		17.2%		17.4%	
16+ Years	42.9%			35.3%		47.9%		57.9%	
Has electricity, computer and internet	77.1%			67.0%		85.7%		95.0%	
Sector:									
Agriculture	5.1%			5.1%		5.2%		5.1%	
Livestock	1.9%			2.2%		1.7%		1.4%	
Food Processing	10.2%			8.9%		11.8%		11.3%	
Apparel	7.4%			6.9%		6.4%		10.5%	
Motor vehicles	2.9%			2.0%		3.2%		5.0%	
Pharmaceuticals	2.7%			1.5%		3.2%		5.3%	
Wholesale or retail	14.7%			18.0%		12.4%		8.6%	
Financial services	3.4%			3.2%		4.0%		2.6%	
Land Transport	5.0%			5.0%		5.2%		4.3%	
Health Services	4.4%			2.1%		5.5%		9.5%	
Leather	2.2%			2.2%		1.8%		2.7%	
Other Manufacturing	15.8%			16.1%		15.1%		16.1%	
Other Services	24.4%			26.7%		24.3%		17.6%	

Notes : The table reports summary statistics of establishment-level measures and distribution of establishments, in the overall sample, and by size-groups. The top panel consists of summary statistics calculated using establishment-level sampling weights. The bottom panel reports the unweighted shares of establishments belonging to the various groups.

Table C.14: Average level of technology measures

Business Function	*ANUM*	*NUM*	*MAX*	*MOST*	N_f	Business Function	*ANUM*	*NUM*	*MAX*	*MOST*	N_f
General Business Functions						*Automotive*					
Business Administration	2.1	2.1	3.0	2.5	5	Vehicle Assembly	2.5	2.2	4.2	1.2	6
Production Planning	1.9	1.9	2.7	2.3	5	Body pressing and welding *	2.3	2.0	1.2	1.3	6
Sourcing	1.9	1.9	2.3	1.7	5	Painting *	1.6	1.8	1.1	1.1	4
Marketing	2.0	2.0	2.4	1.8	5	Plastic injection molding *	1.7	1.9	1.2	1.1	4
Sales	2.2	1.9	2.4	1.6	6	Productive assets management	1.8	2.6	3.1	1.9	3
Payment	2.9	2.2	3.7	2.8	7	Fabrication	2.1	1.9	2.2	1.6	6
Quality Control	1.6	1.8	2.1	1.6	4						
Agriculture						*Pharmaceuticals*					
Land Preparation	1.9	2.2	3.6	3.0	4	Facilities	1.6	1.8	3.7	3.2	4
Irrigation	1.9	1.7	3.3	3.0	6	Weighing scale	1.9	2.2	4.3	3.9	4
Weeding	1.9	1.9	2.5	2.0	5	Mixing/Compounding	1.8	1.8	3.2	2.4	5
Harvesting	1.9	1.9	2.8	2.1	5	Encapsulation	1.8	2.1	4.0	3.2	4
Storage	1.8	1.8	2.8	2.4	5	Quality control	1.6	2.3	3.4	2.8	3
Packaging	1.6	1.8	2.2	1.8	4	Packaging	1.5	2.0	3.4	2.8	3
						Fabrication	2.0	1.8	2.3	1.9	6
Livestock						*Wholesale and Retail*					
Breeding	2.0	2.4	3.1	2.5	4	Customer service	2.3	2.3	2.5	1.5	5
Feeding	3.0	2.3	3.5	2.7	7	Pricing	1.9	1.9	2.3	1.7	5
Animal healthcare	3.2	3.2	4.3	2.9	5	Merchandising	1.6	1.8	2.2	1.6	4
Herd management	2.0	1.7	2.0	1.3	7	Inventory	1.7	1.7	2.4	2.0	5
Transport of livestock	2.0	2.4	3.7	2.8	4	Advertisement	2.5	2.2	3.0	2.3	6
Food Processing						*Financial Services*					
Input testing	1.7	2.0	2.4	1.6	4	Customer service	3.3	4.1	4.7	2.1	4
Mixing/cooking	2.2	2.6	3.2	2.2	4	Avoid fraud	3.4	2.9	3.6	1.7	6
Anti-bacterial	1.9	2.2	2.8	2.1	4	Loan applications	2.7	3.3	3.8	1.7	4
Packaging	1.8	2.1	2.5	1.9	4	Credit applications	2.3	2.7	3.1	1.8	4
Food storage	1.9	2.2	3.1	2.6	4	Operational support area	2.2	3.4	4.3	3.0	3
Fabrication	1.8	1.6	1.9	1.5	6						
Wearing Apparel						*Transportation*					
Design	1.5	2.0	2.7	2.0	3	Planning	1.6	1.8	2.1	1.6	4
Cutting	2.0	2.0	2.6	2.0	5	Plan execution	1.8	1.8	2.2	1.5	5
Sewing	2.1	2.1	3.0	2.4	5	Monitoring	1.9	1.9	2.4	1.6	5
Finishing	1.7	1.7	2.2	1.7	5	Performance measurement	1.8	1.8	2.2	1.7	5
Fabrication	1.9	1.7	1.9	1.6	6	Maintenance	1.6	1.8	2.3	1.7	4
Leather and Footwear						*Healthcare*					
Design	1.5	2.0	2.5	1.9	3	Infrastructure and Machines **	5.5	2.8	2.8	2.8	11
Cutting	2.2	2.2	2.7	2.0	5	Appointment and Scheduling	2.0	2.3	2.6	1.7	4
Sewing	2.1	2.1	2.6	2.2	5	Patient records management	1.9	2.2	3.0	2.4	4
Finishing	1.5	1.5	1.7	1.4	5	Healthcare management ***	1.2	1.9	2.7	0.0	2
Fabrication	1.7	1.5	2.1	1.9	6	Procedures **	4.4	3.0	3.0	3.0	8
						Other Manufacturing					
						Fabrication	1.9	1.7	2.1	1.7	6

Notes : The table provides BF-level averages for the technology measures as described in Appendix B.1. Averages are constructed using establishment-level sampling weights.
* The technology measures for these BFs are calculated differently. Please see Appendix B.1.1 for more details.
** The question on most-used technology ($MOST_{f,j}$) is not asked for this particular BF.

Table C.15: Percentage of establishments with technology gaps

Business Functions	GAP	Business Functions	GAP
General Business Functions		*Automotive*	
Business Administration	34%	Vehicle assembly	91%
Production Planning	27%	Body pressing and welding	0%
Sourcing	15%	Painting	0%
Marketing	15%	Plastic injection molding	0%
Sales	28%	Productive assets management	3%
Payment	48%	Fabrication	21%
Quality Control	5%		
Agriculture - Crops		*Pharmaceutical*	
Land Preparation	37%	Facilities	7%
Irrigation	48%	Weighing	27%
Pest Control	23%	Compounding	24%
Harvesting	58%	Encapsulation	35%
Storage	11%	Quality Control	12%
Packing	9%	Packaging	15%
		Fabrication	11%
Livestock		*Wholesale and Retails*	
Breeding	13%	Customer Service	13%
Nutrition	60%	Pricing	15%
Animal healthcare	58%	Merchandising	7%
Herd management	67%	Inventory	7%
Transport of Livestock	30%	Advertisement	39%
Food Processing		*Finance*	
Input Test	17%	Customer Service	27%
Mixing Blending Cooking	18%	ID Verification	33%
Anti-bacterial	13%	Loan Application	23%
Packaging	5%	Loan Approval	8%
Food Storage	8%	Operational Support Area	0%
Fabrication	8%		
Wearing Apparel		*Transportation*	
Design	8%	Planning	8%
Cutting	6%	Execution	28%
Sewing	10%	Monitoring	39%
Finishing	9%	Performance Measurement	26%
Fabrication	7%	Maintenance	13%
Leather and Footwear		*Healthcare*	
Design	2%	Infrastructure and Machines	81%
Cutting	13%	Scheduling Appointments	5%
Sewing	2%	Management of Patient Records	5%
Finishing	4%	Healthcare Management	0%
Fabrication	3%	Procedures	51%
		Other Manufacturing	
		Fabrication	22%

Notes: The table reports the percentage of establishments that experience sophistication gaps for all business functions. The percentages are calculated using sampling weights.

Table C.16: Variance Decomposition of $D_{f,j}$

Variance of $D_{f,j}$	0.23
Contribution of	
Establishment FE	29%
BF FE	6%

Notes : $D_{f,j}$ is a binary variable that takes value 1 if $MAX_{f,j} > MOST_{f,j}$ and 0 otherwise, provided $NUM_{f,j} > 1$. The bottom panel reports the R^2s of the regressions where $D_{f,j}$ is regressed separately on Establishment and BF FE, respectively. Variance and regression estimates are calculated using establishment-level sampling weights.

Table C.17: Variance Decomposition of MAX and MOST

	$MAX_{f,j}$	$MOST_{f,j}$
Variance	1.60	1.24
Contribution of		
BF FE	16%	17%
Estab. FE	39%	32%
BF and Estab. FE	53%	48%
	MAX_j	$MOST_j$
Variance	0.62	0.41
Contribution of		
Country FE	23%	18%
2-Dig. Sector FE	7%	4%

Notes : The table reports the variance decomposition of both, the BF-level and establishment level measures of MAX and MOST. The contribution reports the R^2s of the regressions where the corresponding MAX or MOST measure is regressed separately on the various fixed-effects. Variance and regression estimates are calculated using establishment-level sampling weights.

Table C.18: Descriptive Statistics of Technology Measures - at BF Level

	Mean	SD	p10	p50	p90
$ANUM_{f,j}$	2.06	1.10	1.00	2.00	4.00
$NUM_{f,j}$	2.00	1.00	1.00	2.00	3.40
$MAX_{f,j}$	2.65	1.27	1.00	2.33	5.00
$MOST_{f,j}$	2.01	1.12	1.00	2.00	4.00

Notes : The table provides BF-level summary statistics for the technology measures as described in Appendix B.1. Statistics are constructed using establishment-level sampling weights.

Table C.19: $MAX_{f,j} > MOST_{f,j}$ conditional on $NUM_{f,j} > 1$ for GBFs

		$MAX_{f,j}$ in			
	Overall	[1,2]	(2,3]	(3,4]	(4,5]
$\Pr(D_{fj} = 1 \mid NUM_{fj} > 1)$	62.6%	54.8%	68.2%	58.5%	69.8%

Notes : This table reports, for the GBFs, the probability of $D_{f,j} = 1$ (i.e. $MAX_{f,j} > MOST_{f,j}$), conditional on the establishment using more than one technology in the BF ($NUM_{f,j} > 1$). Columns 2-5 additionally condition on the value of $MAX_{f,j}$. Calculations are made using establishment-level sampling weights

Table C.20: $MAX_{f,j} > MOST_{f,j}$ conditional on $NUM_{f,j} > 1$ for SSBFs

		$MAX_{f,j}$ in			
	Overall	[1,2]	(2,3]	(3,4]	(4,5]
$\Pr(D_{fj} = 1 \mid NUM_{fj} > 1)$	62.6%	49.4%	63.2%	67.3%	78.5%

Notes : This table reports, for the SSBFs, the probability of $D_{f,j} = 1$ (i.e. $MAX_{f,j} > MOST_{f,j}$), conditional on the establishment using more than one technology in the BF ($NUM_{f,j} > 1$). Columns 2-5 additionally condition on the value of $MAX_{f,j}$. Calculations are made using establishment-level sampling weights

Table C.21: Relationship between technology measures for GBFs

	$MOST_{f,j}$	$MAX_{f,j}$	$MOST_{f,j}$
$MAX_{f,j}$	0.55***		
	(0.01)		
$NUM_{f,j}$		0.86***	0.26***
		(0.01)	(0.01)
N	135063	135063	135063
R-squared	0.68	0.76	0.54
BF FE	Y	Y	Y
Firm FE	Y	Y	Y
Variation Explained	0.34	0.45	0.05

Notes : This table reports the regression estimates, for GBFs, of specifications 4, 5, and 6. To compute the last row, we first residualize the dependent and independent variables by regressing them on the fixed effects, and then we regress the residuals of the dependent on those of the independent. The reported figure is the corresponding R^2. Regressions are estimated using establishment-level sampling weights. Standard errors are clustered at the establishment level. *, ** and *** denote 10%, 5% and 1% significance respectively. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table C.22: Relationship between technology measures for SSBFs

	$MOST_{f,j}$	$MAX_{f,j}$	$MOST_{f,j}$
$MAX_{f,j}$	0.53***		
	(0.02)		
$NUM_{f,j}$		0.79***	0.19***
		(0.02)	(0.02)
N	48841	48841	48841
R-squared	0.70	0.79	0.56
BF FE	Y	Y	Y
Firm FE	Y	Y	Y
Variation Explained	0.33	0.47	0.03

Notes : This table reports the regression estimates, for SSBFs, of specifications 4, 5, and 6. To compute the last row, we first residualize the dependent and independent variables by regressing them on the fixed effects, and then we regress the residuals of the dependent on those of the independent. The reported figure is the corresponding R^2. Regressions are estimated using establishment-level sampling weights. Standard errors are clustered at the establishment level. *, ** and *** denote 10%, 5% and 1% significance respectively. *, ** and *** denote 10%, 5% and 1% significance respectively.

C.2 Scopes and Limits

Table C.24: Variance Decomposition of Specialization, Insourcing and Outsourcing

	$D^{SP}_{f,j}$	$D^{O}_{f,j}$	$D^{IN}_{f,j}$
Variance	0.15	0.05	0.09
Contribution of			
BF FE	10%	6%	7%
Estab. FE	46%	48%	61%
BF and Estab. FE	52%	51%	65%
	SP_j	O_j	IN_j
Variance	0.15	0.05	0.02
Contribution of			
Country FE	4%	2%	7%
Sector FE	8%	9%	4%

Notes : The table reports the variance decomposition of both, the BF-level and establishment level measures of Specialization, Insourcing, and Outsourcing. The contribution reports the R^2s of the regressions where the corresponding Specialization, Insourcing, or Outsourcing measure is regressed separately on the various fixed-effects. Variance and regression estimates are calculated using establishment-level sampling weights.

Table C.23: Technological Sophistication and Establishment Characteristics

	(1) $MOST_{f,j}$	(2) $MAX_{f,j}$
Size: Medium	0.22***	0.31***
	(0.01)	(0.01)
Size: Large	0.49***	0.68***
	(0.01)	(0.01)
Age $<$ 5	0.14***	0.06***
	(0.01)	(0.01)
Age: 6 to 10	0.16***	0.05***
	(0.01)	(0.01)
Age: 11 to 15	0.07***	0.03***
	(0.01)	(0.01)
Foreign owned	0.27***	0.30***
	(0.01)	(0.01)
Exporter	0.23***	0.29***
	(0.01)	(0.01)
Multi-establishment	0.24***	0.37***
	(0.01)	(0.01)
N	172733	175386
R-squared	0.26	0.30
1-Dig. Sector FE	Yes	Yes
Country FE	Yes	Yes
BF FE	Yes	Yes

Notes : Estimates of $MAX_{f,j}$ and $MOST_{f,j}$ on establishment characteristics using establishment-level sampling weights. The base categories are Size: Small, and Age: $\leq$ 5 Years. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table C.25: Country Marginal Effects on SP_j, O_j, and IN_j

Country	SP_j	O_j	IN_j
Korea	0.27	0.05	0.26
Poland	0.16	0.09	0.03
Croatia	0.19	0.03	
Chile	0.12	0.03	
Brazil	0.08	0.05	0.08
Georgia	0.17	0.02	0.05
Vietnam	0.15	0.02	0.01
India	0.22	0.07	0.04
Ghana	0.14	0.02	0.08
Bangladesh	0.14	0.04	0.06
Kenya	0.14	0.04	0.06
Cambodia	0.05	0.03	
Senegal	0.31	0.04	0.03
Ethiopia	0.17	0.04	–0.02
BurkinaFaso	0.39	0.07	
Corr - lnGDP	–0.15	0.16	0.63**
Regression Coefficients:			
lnGDP	–1.10**	–0.18	0.05**
$lnGDP^2$	0.06**	0.01	

Notes : Country marginal effects of SP_j, O_j, and IN_j, estimated by regressing SP_j, O_j, and IN_j on country and 2-Dig. sector FE, using establishment-level sampling weights. *lnGDP* is the log per-capita GDP from Penn World Tables (2019). *Corr* denotes the correlation of the column with *lnGDP*. Regression coefficients are estimates obtained by regressing the column on *lnGDP* and $lnGDP^2$ for SP_j and O_j, and just *lnGDP* for IN_j. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table C.26: Average frequency of ANY_j^{SP}, ANY_j^{IH}, ANY_j^{O}, and ANY_j^{IN}

		Relevant		
				Multi-establishment
Sector	Specialization	Inhouse	Outsourcing	Insourcing
Agriculture	0.47	0.99	0.13	0.24
Livestock	0.48	1.00	0.30	0.16
Food Processing	0.33	1.00	0.10	0.04
Apparel	0.31	0.96	0.17	0.10
Motor vehicles	0.76	0.91	0.46	0.33
Pharmaceuticals	0.82	1.00	0.10	0.16
Wholesale or retail	0.55	0.84	0.46	0.16
Financial services	0.48	0.99	0.12	0.27
Land transport	0.36	1.00	0.16	0.10
Health services	0.20	0.98	0.11	0.17
Leather	0.45	0.97	0.35	0.45
Overall	0.47	0.97	0.22	0.20

Notes : See definitions in Appendix B.2. Averages are computed using sampling weights.

Table C.27: Drivers of ANY_j^{SP}, ANY_j^{O}, and ANY_j^{IN}

	ANY_j^{SP}		ANY_j^{O}	ANY_j^{IN}
	(1)	(2)	(3)	(4)
$\overline{S}_j$	0.11**	0.28***	0.51***	0.53***
	(0.05)	(0.07)	(0.03)	(0.09)
$\overline{S}_j^2$	-0.03***	-0.07***	-0.09***	-0.09***
	(0.01)	(0.01)	(0.01)	(0.02)
$MAX_j - MOST_j$	-0.08***	-0.10***	-0.08***	-0.11***
	(0.01)	(0.01)	(0.01)	(0.02)
Age : 6-10 Years	-0.05***	-0.06***	-0.06***	0.00
	(0.01)	(0.02)	(0.01)	(0.03)
Age : 11-15 Years	-0.08***	-0.06***	-0.07***	0.00
	(0.01)	(0.02)	(0.01)	(0.03)
Age : 16+ Years	-0.05***	-0.07***	-0.09***	-0.10***
	(0.01)	(0.02)	(0.01)	(0.03)
Size : Medium	-0.01	-0.01	-0.03***	-0.01
	(0.01)	(0.02)	(0.01)	(0.02)
Size : Large	0.04	0.13***	-0.06***	0.07**
	(0.02)	(0.03)	(0.02)	(0.03)
Multi-establishment	0.04***	0.01	-0.02*	
	(0.01)	(0.02)	(0.01)	
Export	-0.05***	-0.03	0.01	-0.00
	(0.01)	(0.03)	(0.01)	(0.02)
Foreign owned	-0.00	-0.01	0.05***	0.00
	(0.02)	(0.02)	(0.01)	(0.03)
% of main product in sales		0.08***		
		(0.02)		
N	11627	6465	11214	2319
R-squared	0.12	0.16	0.10	0.15
2-Dig. Sector FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes

Notes : Estimates from Specification 7 with the ANY measures as dependent variables instead. The variables ANY_j^{SP}, ANY_j^{O} , and ANY_j^{IN} are defined as in Appendix B.2. The base categories are Size: Small, and Age: $\leq$ 5 Years. Establishments weighted by sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table C.28: Average frequency of $D^{SP}_{f,j}$, $D^{IH}_{f,j}$, $D^{O}_{f,j}$, and $D^{IN}_{f,j}$ (% of establishments)

		Relevant					Relevant		
				Multi-estab.					Multi-estab.
Business Function	$D^{SP}_{f,j}$	$D^{IH}_{f,j}$	$D^{O}_{f,j}$	$D^{IN}_{f,j}$	Business Function	$D^{SP}_{f,j}$	$D^{IH}_{f,j}$	$D^{O}_{f,j}$	$D^{IN}_{f,j}$
Agriculture					*Pharmaceuticals*				
Land Preparation	10	96	4	1	Facilities	23	85	8	19
Irrigation *	28	100	0	0	Weighing scale **				
Weeding	12	96	4	3	Mixing/Compounding **				
Harvesting	17	95	4	3	Encapsulation	81	84	13	13
Storage	20	95	4	7	Quality control	4	98	2	0
Packaging	42	84	13	20	Packaging	6	98	1	3
Livestock					*Wholesale and Retail*				
Breeding	36	81	17	5	Customer service	13	96	2	11
Feeding	5	98	2	0	Pricing	16	92	5	19
Animal healthcare	19	82	15	14	Merchandising	18	94	4	10
Herd management	4	99	1	0	Inventory	11	95	3	5
Transport of livestock	26	83	14	14	Advertisement	44	83	10	24
Food Processing					*Financial Services*				
Input testing	19	93	7	2	Customer service	17	99	1	0
Mixing/cooking	7	99	1	0	Avoid fraud	13	90	9	5
Anti-bacterial	17	97	3	1	Loan applications	11	95	4	3
Packaging	6	99	1	1	Credit applications	10	94	4	4
Food storage	9	98	1	1	Operational support area	10	96	3	2
Wearing Apparel					*Transportation*				
Design	26	85	14	7	Planning	17	92	6	16
Cutting	11	93	7	2	Plan execution	10	95	5	1
Sewing	12	93	6	5	Monitoring	16	93	6	4
Finishing	15	90	9	6	Performance measurement	10	96	4	2
					Maintenance	15	91	9	3
Automotive					*Healthcare*				
Vehicle Assembly	65	78	12	33	Infrastructure and Machines **				
Body pressing and welding	54	63	37	1	Appointment and Scheduling	10	99	0	1
Painting	75	47	51	8	Patient records management	3	100	0	0
Plastic injection molding	81	46	52	10	Healthcare management *	42	58	35	44
Productive assets management	45	65	35	0	Procedures **				
					Leather				
					Design	36	70	30	11
					Cutting	46	67	33	4
					Sewing	44	71	29	7
					Finishing	38	79	20	6

Notes : The table provides the average frequency (% of establishments) of the BF level measures of specialization, inhouse, outsourcing and insourcing as defined in Appendix B.2. Averages calculated using establishment-level sampling weights.

* The variables are coded differently for these BFs. Check Appendix B.2.1 for further details.

** The questions on inhouse, insourcing and outsourcing are not asked for these particular BFs.

Table C.29: Country Marginal Effects on $D^{SP}_{f,j}$, $D^{O}_{f,j}$, and $D^{IN}_{f,j}$

Country	$D^{SP}_{f,j}$	$D^{O}_{f,j}$	$D^{IN}_{f,j}$
Korea	0.056	0.002	0.053
Poland	0.014	0.001	0.009
Croatia	0.002	0.001	
Chile	0.006	0.001	
Brazil	0.009	0.001	0.007
Georgia	0.015	0.001	0.002
Vietnam	0.003	0.002	–0.000
India	0.016	0.008	0.014
Ghana	0.015	0.000	0.015
Bangladesh	0.013	0.004	0.011
Kenya	0.020	0.001	0.017
Cambodia	0.023	0.002	
Senegal	0.022	0.003	0.008
Ethiopia	0.003	0.000	–0.003
BurkinaFaso	–0.008	0.002	
Corr - lnGDP	0.366	–0.108	0.544*
Regression Coefficients:			
lnGDP	–0.068	0.014	0.009*
lnGDP^2	0.004	–0.001	

Notes : Country marginal effects of $D^{SP}_{f,j}$, $D^{O}_{f,j}$, and $D^{IN}_{f,j}$, estimated by regressing $D^{SP}_{f,j}$, $D^{O}_{f,j}$, and $D^{IN}_{f,j}$ on country and 2-Dig. sector FE, using establishment-level sampling weights. $lnGDP$ is the log per-capita GDP from Penn World Tables (2019). $Corr$ denotes the correlation of the column with $lnGDP$. Regression coefficients are estimates obtained by regressing the column on $lnGDP$ and $lnGDP^2$ for $D^{SP}_{f,j}$ and $D^{O}_{f,j}$, and just $lnGDP$ for $D^{IN}_{f,j}$. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table C.30: Drivers of $D^{SP}_{f,j}$, $D^{O}_{f,j}$, and $D^{IN}_{f,j}$

	$D^{SP}_{f,j}$		$D^{O}_{f,j}$	$D^{IN}_{f,j}$
	(1)	(2)	(3)	(4)
$\overline{S}_j$	0.015***	-0.021***	0.102***	-0.002
	(0.006)	(0.007)	(0.025)	(0.002)
$\overline{S}^2_j$	-0.003**	0.005***	-0.018***	0.001*
	(0.001)	(0.001)	(0.005)	(0.000)
$MAX_j - MOST_j$	-0.003**	-0.007***	-0.010**	-0.004***
	(0.001)	(0.001)	(0.004)	(0.001)
Age : 1-5 Years	0.009***	-0.000	0.049***	-0.001
	(0.002)	(0.002)	(0.009)	(0.001)
Age : 6-10 Years	0.011***	0.004**	0.054***	-0.002**
	(0.002)	(0.002)	(0.009)	(0.001)
Age : 11-15 Years	-0.000	0.002	-0.010	-0.000
	(0.002)	(0.002)	(0.008)	(0.001)
Size : Medium	-0.003**	0.001	-0.008	-0.001
	(0.002)	(0.002)	(0.006)	(0.001)
Size : Large	-0.006**	0.007**	-0.001	-0.004***
	(0.003)	(0.003)	(0.009)	(0.001)
Multi-establishment	0.055***	0.031***		0.003***
	(0.002)	(0.002)		(0.001)
Export	-0.003	-0.003	-0.016**	-0.000
	(0.002)	(0.003)	(0.007)	(0.001)
Foreign owned	-0.001	0.005**	-0.019**	0.003***
	(0.002)	(0.002)	(0.008)	(0.001)
% of main product in sales		0.005**		
		(0.002)		
N	41203	22661	9187	40907
R-squared	0.11	0.14	0.11	0.15
2-Dig. Sector FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
BF FE	Yes	Yes	Yes	Yes

Notes : Estimates from Specification 7 with the BF-level measures as dependent variables instead. The variables $D^{SP}_{f,j}$, $D^{O}_{f,j}$, and $D^{IN}_{f,j}$ are defined as in Appendix B.2. The base categories are Size: Small, and Age: $\leq$ 5 Years. Establishments weighted by sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table C.31: Productivity and Other Technology Measures

	log(sales per worker)	
	(1)	(2)
K_j	0.247***	0.231***
	(0.007)	(0.007)
H_j	0.308***	0.160***
	(0.041)	(0.041)
Computers	0.559***	0.299***
	(0.044)	(0.045)
$\overline{S}_j$		0.464***
		(0.020)
Constant	6.759***	6.097***
	(0.176)	(0.175)
N	12966	12966
R-squared	0.385	0.410
2-Dig. Sector FE	Yes	Yes
Country FE	Yes	Yes

Notes : Regression estimates of productivity on different technology measures. Computers is 1 if establishment has computers, electricity and internet, and 0 otherwise. *, ** and *** denote 10%, 5% and 1% significance respectively.

C.3 Productivity and Sophistication

Table C.32: Productivity and Vertical Dimensions

	log(sales per worker)			
	(1)	(2)	(3)	(4)
K_j	0.31***	0.32***	0.31***	0.31***
	(0.01)	(0.01)	(0.01)	(0.01)
H_j	0.10*	0.12**	0.05	0.05
	(0.06)	(0.06)	(0.06)	(0.06)
$\overline{S}_j$	0.34***	0.33***	0.36***	0.36***
	(0.03)	(0.03)	(0.03)	(0.03)
SP_j	-0.09			
	(0.06)			
ANY_j^{SP}		-0.07**		
		(0.03)		
$Sourcing_j$			-0.11	
			(0.08)	
$ANY_j^{Sourcing}$				-0.08*
				(0.04)
N	7776	7885	7646	7603
R-squared	0.42	0.42	0.42	0.42
2 Dig. Sector FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes

Notes : The table reports relationship between productivity and vertical dimensions of the grid. SP_j and ANY_j^{SP} are defined as in Appendix B.2. $Sourcing_j$ refers to the fraction of relevant BFs that are either insourced or outsourced. $ANY_j^{Sourcing}$ is a binary variable that takes value 1 for the establishment if the establishment has atleast one BF that is sourced (insourced or outsourced), and 0 otherwise (provided there is atleast one relevant BF). Regressions are calculated using establishment-level sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table C.33: Productivity and Vertical Dimensions - MAX v/s MOST

	log(sales per worker)			
	(1)	(2)	(3)	(4)
K_j	0.31***	0.32***	0.31***	0.31***
	(0.01)	(0.01)	(0.01)	(0.01)
H_j	0.15***	0.16***	0.10*	0.10*
	(0.06)	(0.06)	(0.06)	(0.06)
MAX_j	-0.05	-0.05	-0.03	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)
$MOST_j$	0.45***	0.43***	0.44***	0.45***
	(0.04)	(0.04)	(0.04)	(0.04)
SP_j	-0.11*			
	(0.06)			
ANY_j^{SP}		-0.09***		
		(0.03)		
$Sourcing_j$			-0.10	
			(0.08)	
$ANY_j^{Sourcing}$				-0.09**
				(0.04)
N	7776	7885	7646	7603
R-squared	0.43	0.43	0.43	0.43
2 Dig. Sector FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes

Notes : The table reports relationship between productivity and vertical dimensions of the grid. SP_j and ANY_j^{SP} are defined as in Appendix B.2. $Sourcing_j$ refers to the fraction of relevant BFs that are either insourced or outsourced. $ANY_j^{Sourcing}$ is a binary variable that takes value 1 for the establishment if the establishment has atleast one BF that is sourced (insourced or outsourced), and 0 otherwise (provided there is atleast one relevant BF). Regressions are calculated using establishment-level sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table C.34: Development Accounting Contributions and Distribution of Residuals

Sector	Country FE	Contribution	Sales Residual		$\overline{S}_j$ Residual	
			p10	p90	p10	p90
Overall	Y	0.24	–1.38	1.45	–0.63	0.75
	N	0.31	–1.55	1.63	–0.71	0.84
	Sector X Country	0.25	–1.34	1.41	–0.64	0.74
Agriculture	Y	0.33	–0.99	2.19	–0.75	0.86
	N	0.50	–0.78	2.61	–0.71	0.93
Manufacturing	Y	0.26	–1.07	1.90	–0.58	0.84
	N	0.30	–1.27	2.31	–0.70	0.93
Services	Y	0.24	–1.49	1.35	–0.63	0.78
	N	0.28	–1.66	1.55	–0.68	0.86

Notes : The table reports the contribution of $\overline{S}_j$ in explaining variation in Sales per worker. To calculate the contributions, we first calculate the residuals of Sales per worker and $\overline{S}_j$ by separately regressing them on Sector and (with or without) Country FE (or Sector X Country FE). Columns 4-7 report the distribution of the calculated residuals. Then using the coefficients of $\overline{S}_j$ in the corresponding productivity regressions (columns 1-3 and 6-11) in Table 14, we calculate the contribution in each case using the following formula : $\frac{p90_{\overline{S}_j} - p10_{\overline{S}_j}}{p90_{Sales} - p10_{Sales}} * \beta_{\overline{S}_j}$, where $\beta_{\overline{S}_j}$ is the corresponding coefficient of $\overline{S}_j$ in Table 14.

Table C.35: $MOST_j$ vs. narrow establishment-level technology measures

	$MOST_j$								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Computers	0.75***								
	(0.01)								
ERP		0.74***							
		(0.01)							
Robots			0.79***						
			(0.04)						
$\overline{S}_{BusAdmin,j}$				0.37***					
				(0.00)					
$\overline{S}_{Payments,j}$					0.34***				
					(0.01)				
$\overline{S}_{Fabrication,j}$						0.43***			
						(0.01)			
$\overline{S}_{GBF,j}$							0.84***		
							(0.00)		
$\overline{S}_{SSBF,j}$								0.10***	
								(0.00)	
$\overline{S}_{GBF,j}$ * Agriculture									0.54***
									(0.03)
$\overline{S}_{GBF,j}$ * Manufacturing									0.73***
									(0.01)
$\overline{S}_{GBF,j}$ * Services									0.84***
									(0.00)
$\overline{S}_{SSBF,j}$ * Agriculture									0.35***
									(0.02)
$\overline{S}_{SSBF,j}$ * Manufacturing									0.17***
									(0.01)
$\overline{S}_{SSBF,j}$ * Services									-0.03***
									(0.00)
N	20407	19530	7238	20383	20566	6908	20995	18316	18310
R-squared	0.14	0.20	0.05	0.49	0.18	0.24	0.80	0.03	0.80
1-Dig. Sector FE	No	No	No	No	No	No	Yes	Yes	Yes

Notes : Estimates from running the regressions as in Table 13, with $MOST_j$ as the independent variable instead. All regressions use establishment-level sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table C.36: MAX_j vs. narrow establishment-level technology measures

	MAX_j								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Computers	1.03***								
	(0.02)								
ERP		1.14***							
		(0.01)							
Robots			1.14***						
			(0.05)						
$\overline{S}_{BusAdmin,j}$				0.44***					
				(0.00)					
$\overline{S}_{Payments,j}$					0.42***				
					(0.01)				
$\overline{S}_{Fabrication,j}$						0.54***			
						(0.01)			
$\overline{S}_{GBF,j}$							1.04***		
							(0.00)		
$\overline{S}_{SSBF,j}$								0.20***	
								(0.00)	
$\overline{S}_{GBF,j}$ * Agriculture									0.73***
									(0.03)
$\overline{S}_{GBF,j}$ * Manufacturing									0.82***
									(0.01)
$\overline{S}_{GBF,j}$ * Services									1.06***
									(0.00)
$\overline{S}_{SSBF,j}$ * Agriculture									0.33***
									(0.02)
$\overline{S}_{SSBF,j}$ * Manufacturing									0.26***
									(0.01)
$\overline{S}_{SSBF,j}$ * Services									0.04***
									(0.00)
N	20407	19530	7238	20383	20566	6908	20995	18316	18310
R-squared	0.18	0.33	0.07	0.48	0.18	0.30	0.84	0.08	0.85
1-Dig. Sector FE	No	No	No	No	No	No	Yes	Yes	Yes

Notes : Estimates from running the regressions as in Table 13, with MAX_j as the independent variable instead. All regressions use establishment-level sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

D Detailed acknowledgments

The preparation of the Firm-level Adoption of Technology survey questionnaire involved the contribution of several sector experts within and outside the World Bank.

First, we would like to thank the following World Bank Group colleagues: Erick C.M. Fernandes (Lead Agriculture Specialist), Holger A. Kray (Manager, Agriculture), Michael Morris (Lead Agriculture Economist), Madhur Gautam (Lead Agriculture Economist), Robert Townsend (Lead Agriculture Economist), Ashesh Prasann (Agriculture Economist), Aparajita Goyal (Senior Economist, specialized in Agriculture), Harish Natarajan (Lead Financial Sector Specialist), Erik Feyen (Lead Financial Sector Economist), Laurent Gonnet (Lead Financial Sector Specialist), Arturo Ardila Gomez (Lead Transport Economist), Victor A Aragones (Senior Transport Economist), Edson Correia Araujo (Senior Health Specialist), Irina A. Nikolic (Senior Health Specialist), Brendan Michael Dack (Chief Industry Specialist at IFC), Emiliano Duch (Lead Private Sector Specialist), Blair Edward Lapres (Economist), Kazimir Luka Bacic (Operations officer), Justin Hill (Senior Private Sector Specialist), Etienne Raffi Kechichian (Senior Private Sector Specialist), Justin Yap (Private Sector Specialist), Austin Kilroy (Senior Economist).

Similarly, we would like to thank several external experts. From Embrapa-Brazil, we would like to thank Alexandre Costa Varella (Head of Embrapa research unit in the South region, expert on livestock), Flávio Dessaune Tardin (Researcher, Maize and Sorghum), Edison Ulisses Ramos Junior (Researcher, Soybean), Isabela Volpi Furtini (Researcher, Rice and Beans), Alberto Duarte Vilarinhos (Researcher, Cassava and Fruits), Carlos Estevão Leite Cardoso (Researcher, Cassava and Fruits, Technology transfer), and other participants of the internal seminars to validate the sector-specific questionnaire for agriculture and livestock. We also would like to thank a group of senior consultants from Patina Solutions, who contributed with the preparation and validation exercise for several sectors, including Daren Samuels (Consultant, Manufacturing), Christ Baughman (Consultant, Food Processing), Sandra Aris (Consultant, Wearing Apparel and Lather), Steve Zebovitz (Consultant, Pharmaceutical sector), Shelly Wolfram (Consultant, Retail and Wholesale), James M. Keding (Consultant, Transport and Logistics), as well as Justin Barners (Consultant, Automotive). Finally, we would like to thank Sudha Jayaraman (Associate Professor, Department of Surgery, Virginia Commonwealth University), Christina Kozycki (Infectious Disease Fellow, NIH), Jonathan Skinner (Health Economist, Dartmouth College), and Elizabeth Krebs (Assistant Professor of Emergency Medicine at the Jefferson Center for Injury Research and Prevention), for their comments and feedback on health services.

We also would like to thank the National Statistics Agencies and local partner institutions

that provided access to the sample frame and/or support the implementation of the survey. They include: Statistics Korea, Statistics Poland, National Statistics and Demography of Senegal (ANSD), the Federation of Industries of the State of Ceará (FIEC), and the General Statistics Office of Vietnam (GSO), the Kenya National Bureau of Statistics (KNBS), the Bangladesh Bureau of Statistics, and the Central Statistics Office of India.