NBER WORKING PAPER SERIES

TECHNOLOGY SOPHISTICATION ACROSS ESTABLISHMENTS

Diego A. Comin
Xavier Cirera
Marcio Cruz

Working Paper 33358
http://www.nber.org/papers/w33358

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2025

We thank David Baqaee, Jan DeLoecker, Mary Hallward-Driemeier, Apoorv Gupta, John Haltiwanger, Maurice Kugler, Bill Maloney, Martha Martinez Licetti, Paolo Mauro, Denis Medvedev, Dani Rodrik, Giacomo Ponzetto, Tommaso Porzio, 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, Antonio Martins Neto, Soumya Agrawal, and Adrian Forest 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.(2020), Cirera et al. (2021) and Cirera, Comin and Cruz (2024). 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.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2025 by Diego A. Comin, Xavier Cirera, and Marcio Cruz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Technology Sophistication Across Establishments
Diego A. Comin, Xavier Cirera, and Marcio Cruz
NBER Working Paper No. 33358
January 2025
JEL No. O1, O3, O4

ABSTRACT

This paper examines technology sophistication in establishments. To comprehensively measure technology sophistication, we create a grid that covers key business functions and the technologies used to conduct them. Analyzing data from over 21,000 establishments in 15 countries, we find that the most widely used technology is usually not the most sophisticated available in the business function. There is significant variation in technology sophistication across and within countries, explaining 31% of productivity dispersion and over half of the agricultural productivity gap. The sophistication of widely used technologies is more relevant for productivity than the most advanced technologies. More sophisticated technologies are appropriate for both developed and developing countries.

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

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

1 Introduction

Technology is central to some of the most fundamental economic questions. Yet, our understanding of these issues often depends on indirect and limited measures. A long tradition, dating back to Ryan and Gross (1943) and Griliches (1957), has characterized technology in establishments by the presence of a few (typically one) advanced technologies. This approach faces several limitations. First, the number of technologies is very small compared to those used in an establishment. Additionally, the tasks where establishments use these technologies are neither comprehensive or representative of the business functions conducted in an establishment. Second, measures based on the presence of advanced technologies do not provide information on how establishments without them produce. In particular, we do not know how sophisticated the technologies used are relative to the frontier. This concern is particularly relevant in developing countries where advanced technologies are less widely diffused. Third, traditional measures do not capture how intensively a technology is used, which is crucial to explain income divergence across countries (Comin and Mestieri, 2018). This omission limits our understanding of whether establishments predominantly use the most sophisticated technologies they have adopted and the importance for productivity of these technologies relative to the most widely used technologies.[1]

In this paper, we develop a new approach to directly and comprehensively measure the sophistication of technologies used in establishments. Our first step is to create a two-dimensional grid structure, which we refer to as 'the grid'. Following the task-based production function approach (Zeira, 1998; Grossman and Rossi-Hansberg, 2008; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018), the horizontal dimension of the grid covers the key tasks an establishment conducts, grouped into broader categories that we call business functions (BF). To this, we add a vertical dimension that represents the range of technologies that can be used to perform the key tasks in each business function. The grid encompasses

[1] Since the classic work of Ryan and Gross (1943) and Griliches (1957) on hybrid corn, many have applied this approach to measuring technology in establishments in other sectors. For example, Davies (1979) studies the diffusion of 26 different manufacturing technologies, each 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 the internet. Other efforts include the Survey of Manufacturing Technology by the Census Bureau discontinued after 1993, which covers 17 specific technologies, including numerically-controlled machines, computer-aided design or engineering technologies, programmable controllers and local area networks (see Dunne (1994)); or the Canadian Survey of Advanced Technologies with 41 and 50 technologies, depending on the round (see for example Boothby, Dufour and Tang (2010)). More recently, the Advanced Business survey, Acemoglu et al. (2022), also administered by the US Census Bureau and that focused on five generic, frontier technologies: AI, robotics, dedicated equipment, specialized software and cloud computing. Unlike the previous studies, the Advanced Business Survey asks for the intensity with which the firm uses these advanced technologies.

63 business functions: seven general business functions (GBF) relevant to all sectors, and 56 sector-specific business functions (SSBF) across 12 sectors (agriculture, livestock, food processing, apparel, leather goods, automotive, pharmaceutical, other manufacturing, wholesale and retail, financial services, land transport services, and health services). 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, the technologies in each business function are ranked according to their sophistication, from the simplest to the most complex which represents the world technology frontier.

We implement the grid in the Firm Adoption of Technology (FAT) survey, administered to over 21,000 establishments that constitute representative samples in 15 countries: 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. FAT collects three types of information. First, it gathers establishment-level data on sales, inputs, education of the workers and the managers, management practices, etc. Second, it records whether each sector-specific business function is conducted in-house. Third, and most relevant, FAT documents the technologies from the grid used by each establishment in each business function and, of these, which one is the most widely used technology.

Using the information from FAT, we develop two measures of technology sophistication at the business function-establishment level: 'MOST' for the most widely used technology, and 'MAX' for the most advanced technology available. We use these measures to study three topics: the use of technology at the business function level, the cross-establishment variation in technology sophistication, and the relationship between technology sophistication and productivity across establishments.

At the business function level, our analysis reveals that the most widely used technology (MOST) is usually not the most sophisticated one available (MAX). The gap between MAX and MOST is persistent, indicating that MAX and MOST represent two distinct and relatively independent processes of technology upgrading within establishments.

By aggregating across all functions of an establishment, we derive establishment-level measures of technology sophistication that inherit the comprehensiveness of the grid. We observe significant variation in technology sophistication across establishments, both across and within countries. The dispersion in technology sophistication across establishments varies considerably across countries, increasing with per-capita income. Technology sophistication is positively associated with establishment size, the human capital of its workers, the quality of management practices, exporter status, multi-national and multi-establishment status, and shows an inverted U-shaped relationship with establishment age.

Examining the relationship between technology sophistication and productivity across establishments, we show that technology sophistication is strongly associated with productivity. Differences in technology sophistication account for 31% of the variation in productivity across establishments. This result holds even when controlling for management quality, human and physical capital, and markups.

The productivity regressions yield three, additional insights. First, MOST is more relevant than MAX for establishment productivity, highlighting the importance of focusing on MOST in technology measurement and modeling. Second, there is significant variation across sectors in the share of productivity dispersion accounted for by technology sophistication. For example, it accounts for 50%, in agriculture but only for 28% in services. Consequently, differences in technology sophistication account for more than half of the agricultural productivity gap between high- vs. low-income economies (Caselli, 2005). Third, we examine whether technology is equally beneficial 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. This elasticity is not smaller in low-income countries, suggesting that the productivity gains from using more sophisticated technologies are not limited to advanced economies.

Our study of technology sophistication is closely related to various lines of research. The patterns documented about the use of technology at the business function level provide an empirical counterpart to the two frameworks used to model technology in production: quality ladders (Aghion and Howitt, 1992) and love-for-variety models (Romer, 1990). These canonical paradigms predict that the best available technology at the business function level (MAX) suffices to characterize the technology sophistication of the business function and drives the establishment's productivity. However, our findings highlight that MOST is not only key to characterizing the sophistication of technologies used in a business function, but is also more strongly associated with productivity across establishments than MAX.

The notions of inter- and intra-firm diffusion are connected to MAX and MOST. Mansfield (1963) introduced the concept of intra-firm diffusion to describe the gradual increase in the use of a technology (e.g., diesel locomotives) within a firm after its adoption. The small body of literature following Mansfield has studied the intra-firm diffusion of a few technologies, in a few countries, such as numerically controlled machines in UK metalworking (Battisti and Stoneman, 2003) and e-commerce in the UK and Switzerland (Battisti et al., 2007; Hollenstein and Woerter, 2008). These studies have produced mixed results on whether intra- and inter-firm diffusion follow similar processes and whether earlier (less-sophisticated) technologies remain the most widely used for a long time.

Our broader study establishes general patterns about the gap between MAX and MOST. Additionally, with the comprehensive coverage of technologies and productivity measurement at the establishment level, we reveal a novel finding: the different associations between productivity and the MAX and MOST measures of sophistication.

There is a long tradition studying the relationship between the adoption of advanced technologies by an establishment and productivity.[2] Studies in this literature typically consider a limited number of technologies. For example, Acemoglu et al. (2022) find an association between productivity and the number of advanced technologies a firm adopts (see footnote 1 for a list of the five technologies considered). Our analysis of the relationship between technology sophistication and productivity builds on this literature. In addition to the greater number of technologies considered, our exploration extends previous work in three directions. First, the use of a grid that provides a dense coverage of business function and ensures that the technologies considered are representative of the main tasks conducted in the establishment. In particular, they do not just reflect generic tasks that apply to establishments in a wide range of sectors but they also include a large number of tasks that are specific to individual sectors. Second, the estimates found in the literature (e,g,. Acemoglu et al. (2022)) are in line with the coefficient for MAX in the productivity regression. However, the literature has not included measures of MOST in the productivity regressions. A key novel finding in our analysis is that the association of productivity and technology sophistication is much stronger with MOST than with MAX. This cross-establishment finding is consistent with the findings in Comin and Mestieri (2018) who study the relationship between productivity growth and the intensity of use of technologies across countries and over time. Third, most studies of productivity and technology across establishments are limited to a single country. Even though our 15 country-sample is far from representative of the 200+ countries in the world, it suffices to demonstrate that the strong association between technology sophistication and productivity holds both within and between countries and that the within-country association does not differ between developed and developing economies.

There are clear methodological and conceptual parallels between our effort to measure and study technology sophistication and the seminal work by Bloom and Van Reenen (2007) on management practices. As with technology, there is a long tradition of documenting specific management practices in a limited number of companies. The groundbreaking studies by Bloom and Van Reenen (2007) and Bloom et al. (2019) have greatly extended this scope by

[2]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.

measuring the quality of management practices across 18 dimensions related to operations, planning, monitoring, and human resources, covering thousands of firms in many countries.

Similar to FAT, data on management practices is collected via firm surveys. Experts rank practices based on their quality, and an establishment-level score is constructed to study the drivers of management practices and their association with productivity. We perform a similar analysis for technology sophistication. Beyond the similarities in measurement methods, Bloom, Sadun and Reenen (2012) have hypothesized that technology sophistication and management practices are complementary. We explore the complementarity of technology and management in the context of productivity regression across establishments, finding supporting evidence.

The rest of this paper is organized as follows. Section 2 introduces the FAT survey, and describes various validation exercises of the sophistication rankings, and the data collected. Section 3 presents the technology sophistication measures and illustrates key insights with examples from specific establishments, and sectors in FAT. Section 4 explores the use of technology at the business function level. Section 5 studies technology sophistication across establishments. Section 6 investigates the relationship between technology sophistication and productivity 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 each establishment uses to perform key business functions necessary to operate in its respective sector. In the following sub-sections, we describe the survey design and implementation, relegating further details to section A in the Appendix.

2.1 Structure

The survey is composed of five modules. Module A collects information on the general characteristics of the establishment.[3] 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 financial statements and employment.

[3]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. The survey can be downloaded at the following address (https://dcomin.host.dartmouth.edu/files/FAT_Survey_complete.pdf). The implementation manual which includes all instructions for interviewers, training materials and a full description of the technologies in the grid can be downloaded at (https://dcomin.host.dartmouth.edu/files/Implementation_Manual_TAS_29112023.pdf).

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, including agriculture (crops and fruits), livestock, food processing, wearing apparel, leather goods and footwear, automotive, pharmaceutical, other manufacturing, wholesale and retail, financial services, land transport services, and health services.[4] These sectors have been selected based on their share in aggregate value-added, employment and number of establishments and they cover all three industries (agriculture, manufacturing, and services).

2.2 The Grid

To design Modules B and C, we determined the business functions covered and the list of technologies, from most basic to most sophisticated, 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, 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).

All technologies in the Grid are precisely described so that respondents and enumerators can objectively establish their use. Figure 1 presents the general business functions considered in the survey and the possible technologies that can be used to conduct each of

[4]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).

them. The grid contains 7 GBFs: Business administration, production planning, sourcing and procurement, marketing, sales, payment methods, and quality control. Each function considers between 4 and 7 technologies. For example, an establishment can gather and analyze customer information for marketing purposes using face-to-face conversations, online chats via WhatsApp or the internet, structured customer surveys, customer relationship management (CRM) software to store contact information, interaction history, and communication preferences, or big data analytics and/or artificial intelligence to uncover trends and make informed marketing decisions. Figure 2 presents the grid for one sector-specific module, agriculture. The grid considers six SSBFs for agriculture which include land preparation, irrigation, weeding and pest management, harvesting, storage and packaging. For example, to prepare the land for cultivation, a farm can use manual labor with simple tools such as hand-held hoes, or rakes, animal-aided instruments such as ploughs, equipment manually operated such as tractors, motor tillers, or rotators, or equipment supported by digital technologies such as GPS, software or precision agriculture tools. Section A.1.1 of the appendix reports the grids for all other SSBFs and the implementation manual precisely defines each of the technologies in the grid.

2.3 Ranking of Technology Sophistication

In addition to identifying key business functions and relevant technologies, industry experts ranked the technologies in each function based on their sophistication. More sophisticated technologies can perform a wider variety of tasks, more complex tasks, or perform tasks with greater accuracy and speed. The experts' deliberations and resulting sophistication rankings, shown on the grid, were produced before the survey administration. This approach to ranking technologies resembles the World Management Survey (Bloom and Van Reenen, 2007), which relies on experts to rank management practices according to their quality.

Given the importance of the ranking for our analysis, we evaluated the coherence of the expert rankings through a three-stage validation process implemented in 14 of the 63 business functions on the grid including most of the GBFs and SSBFs in agriculture, food processing apparel, and retail. The three stages are as follows:

1. **Comparison of Key Features**: We compared the technologies in each business function along three dimensions invoked by the experts: functionality, integration, and automation. Functionality refers to the capabilities a technology offers to handle more complex tasks, in a faster way, on a larger scale, with greater accuracy and reliability. Integration reflects a technology's ability to connect and interact seamlessly with other systems by exchanging data and coordinating processes. Automation enables the technology to execute

processes, make decisions, and generate outcomes independently, without human intervention.

2. **Novelty and Cost**: We documented the year of invention and the cost of each technology and studied their correlation with the experts' rankings. Although novelty and cost do not define sophistication, more sophisticated technologies tend to be newer and more expensive.

3. **Large language models (LLMs)**: We conducted two exercises to validate our ranking through LLMs. First, we asked ChatGPT to rank the technologies based on their levels of sophistication. We replicate this exercise following specific definitions of sophistication based on functionality, integration, and automation. Second, we asked ChatGPT to identify a specific task for each of the 14 business functions and estimate the time required to perform the task with each technology.

To collect the information in the first two stages, we relied on multiple sources, including the official description of specific leading brands supplying these technologies. For GBFs we collected information from multiple companies websites, including Microsoft, Google, SAP, Oracle, QuickBooks, IBM, Sage, NetSuite, BambooHR, Trello, Salesforce, Workday, Meta, Qualtrics, Survey Monkey, Amazon, Shopify, LinkedIn, among others. These companies have more than 80% of the global market share for technologies used in business administration, such as standard software (e.g, spreadsheet) and enterprise resource planning (ERP) systems, and a large share of the market across GBFs.[5] In addition, we consulted specialized websites (e.g, tech.co; erpresearch.com; getapp.com) that provide comparisons across these products, totaling more than 50 original sources of information. Similar exercises with multiple sources of information were replicated for SSBFs.

We illustrate the validation methodology using the example of business administration, a GBF that includes finance, accounting, and human resources processes. Table 1 summarizes the three-step validation procedure for each technology in business administration. The least sophisticated technology, handwritten processes, can only perform basic manual administration tasks such as transaction entry, bookkeeping, or employee records handling, without any integration or automation features. Standard software like Microsoft Excel or Google Sheets helps with basic functionality to perform mathematical and statistical operations, including charts, and handle financial account, and HR records. However, it requires manual inputs and knowledge to build specific applications, with limited integration and automation. Mobile apps, such as QuickBook online, are pre-designed to perform these tasks with some

[5] Estimates based on Enlyft dataset (Cirera, Comin and Cruz, 2022). These companies are recognized as key players by various specialized sources estimating market potential for ERP (e.g., Research and Market, Fortune Business Insight), even if there are variations in their market share estimations.

integration and automation features, but with limited scale and customization. Specialized software, such as Oracle Financial, have high functionality, integration, and automation capabilities, within specialized domains. Finally, enterprise resource planning (ERP systems), such as SAP and Oracle NetSuite provide comprehensive functionality with full integration within and across business functions, with a high level of automation. Comparing the features of business administration technologies results in a ranking that matches the grid's sophistication ranking.

More sophisticated technologies in business administration are embodied in more expensive software. For example, standard software such as Microsoft 365 (which includes Microsoft Excel) and Google Sheets costs between \$12 and \$18 per user/month; apps such as Quickbooks Online and BambooHR cost between \$30 and \$200 per user/month; specialized software such as Oracle Financials, Intuit Quickbooks and Workday HCM cost between \$120 and\$600 per user/month; ERP systems such as SAP ERP or Oracle Netsuite cost over \$1700 per user/month. In business administration, there is no clear relationship between technology novelty and our sophistication ranking (e.g., the SAP ERP system was introduced in 1981, while Microsoft Excel was first available in 1985).As shown in section A.2 of the appendix, for most SSBFs, especially in agriculture and manufacturing, we observe a positive and strong association between technology novelty, cost, and our sophistication rankings.

ChatGPT's ranking of the technologies in business administration based on functionality, automation, and integration coincides with the expert ranking. Furthermore, the ranking is robust to variations of the prompts provided to ChatGPT focusing on specific dimensions of technology sophistication (e.g, exclusively based on functionality, integration, or automation). Finally, the estimated time to perform a task (e.g., managing payroll and financial reports) provided by ChatGPT is consistent with the experts' sophistication ranking. For example, it takes roughly 5 hours to conduct the task with handwritten processes, 2 hours with computer and standard software, 1.5 hours with an app, 1 hour with specialized software and 30 minutes with an ERP system. We replicated these exercises for business administration in ChatGPT over 100 iterations, to account for its probabilistic features and potential variation in a typical activity, and the patterns are consistent. The ChatGPT rankings are strongly and positively associated with the experts' rankings, following a similar ranking order of sophistication in all iterations.

Overall, the validation exercise across all 14 business functions supports the experts' ranking of sophistication.[6]

[6] In section A.2 of the appendix, we provide the results for the other 13 business functions where we implement the validation of the expert rankings.

2.4 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 financial statements, workers, and management.

Business functions. The business functions that comprise the horizontal 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.[7] 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. This information helps us assess the relevance of establishment-level measures of technology sophistication based only on the technologies used in functions conducted in-house.

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 3.6% 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.

Other variables. The survey also includes other standard questions about financial statements' information, employment, education of the employees, and education and experience of the manager. The survey collects information on four management practices from MOPS, including the presence of formal incentives, number of key performance indicators (KPIs),

[7] 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.

frequency of KPI review, and time frame of production targets. 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 based on this subset of questions accounts for 90.5% of the cross-establishment variance of the original MOPS z-score for Mexican establishments collected by ENAPROCE.

2.5 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, and Europe), with different levels of income, for which there was access to a high-quality sampling frames. In these countries, we collected data from 21,055 randomly selected establishments from the sampling frames. Table 2 shows the distributions of the sample by country, sector, and size groups and Table C.1 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.

2.5.1 Sampling

Our data is representative for a universe of about 2.1 million establishments. The samples are nationally representative for establishments with 5 or more workers. 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. 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.[8]

[8]Table A.15 provides information about the distribution of firms by country, sector, and size groups within the universe covered by the FAT survey. 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. Section A of the Appendix provides more details on the sampling frame, survey implementation and data collection, and sampling weight.

2.5.2 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 designing and implementing the FAT survey to minimize these errors. Appendix A.5 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 had demonstrable experience in nationally representative firm-level surveys. Third, we followed a standard protocol in which each firm was contacted several times to schedule an interview. Fourth, we mostly used face-to-face or phone interviews, which usually have higher response rates than web-based interviews.[9]

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, and we regularly monitored the data collection process using standard algorithms to analyze the consistency of the data.[10]

[9]These procedures are in line with suggestions of good practice for implementation by (Bloom et al., 2016). We use online surveys only for Georgia and Croatia. In Georgia, we partnered with the National Statistical Office, which resulted in exceptionally high response rate. Face-to-face interviews were not possible during the pandemic. See Table A.16 for the mode and date of data collection in each country.

[10]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 or CATI (Caeyers, Chalmers and De Weerdt, 2012). For Georgia and Croatia, we used Computer Assisted Web Interviewing (CAWI).

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.

2.5.3 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 establishment-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). 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.[11] We conducted three tests to assess potential biases from unit non-response-rates.[12] In each of these exercises, presented in Section A.6 of the Appendix, we find no statistical difference in the number of employees, technological sophistication, wages, and share of workers by skill and education

[11] Table A.17 in the Appendix A provides the response rate by country, defined as the ratio between establishments that responded to the survey and the total number of eligible establishments in the sample for which we attempted to conduct an interview. The response rates were higher when national statistical agencies implemented the survey. Section A.4 of the appendix provides more details on sampling weights.

[12] 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.18 to A.24 in Appendix A.

between establishments in the group that proxies for the response sample and the group of establishments 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 establishments with a short version of the questionnaire. For those establishments, we randomly selected three business functions and asked about the presence of the relevant technologies. We estimated a probit model to assess the likelihood of consistent answers between the original and the back-check interviews, controlling for establishment-level fixed-effects. Reporting the use of a technology in the back-check interview is associated with 80.6% of the 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 survey. These estimates do not differ between establishments of different sizes.[13]

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). We focus on variables related to establishment size, productivity and technology. Table A.24 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. In the Brazil matched establishments, we find a strong correlation between FAT measures of log value-added per worker and the log of average wages from RAIS (See Table A.23). 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).[14] 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

[13] 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. 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.

[14] 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 on salaries for all formal workers in Brazil. The FAT survey asks about sales and the number of employees for two periods. The most recent year for which the information is available (i.e. the year before the implementation of the survey) and two years before that. For Korea, these reference years are 2019 and 2017.

association between the book value of machinery and equipment in KED and the establishment technology 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, which are the more sophisticated technologies in each BF and are marked in bold in Appendix A.1. 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.[15]

These ex-post checks further reassure us about the soundness of the survey design, the data collection process, and the accuracy of responses.

3 Measures of Technology Sophistication

We next introduce measures of technology at the business function and establishment levels, constructed using information collected by the FAT survey. Before analyzing these measures, we illustrate the granularity of the grid and how these measures can characterize the sophistication of technology used by establishments, with examples from FAT.

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 ($SG_{f,j}$) 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

[15]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).

rank $\tau + p$, for $1 \leq p < k$. $SG_{f,j}$ is 0 when there are no gaps and at least two technologies are used in the function.[16]

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, $r_f \in 1, 2, ..., R_f$.[17] 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 define a scaled measure of the number of technologies used in a business function ($NUM_{f,j}$).[18]

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 an 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 used in a given business function. This technology may not be new to the establishment, but it is new to the business function of the establishment.

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

[17] 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.

[18] 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 the business function f.

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 when a new establishment's most widely used technology in the business function is more sophisticated than the previous one. This new technology may be entirely new to the business function or an existing technology whose use has been expanded. Therefore, $MOST_{f,j}$ is more related to Mansfield (1963)'s concept of technology diffusion within the firm, specifically within the business function, rather than to innovation.

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

Before studying the general patterns of technology use in establishments, it is useful to become familiar with the grid and the measures of technology sophistication by exploring some examples from FAT.

To begin appreciating the level of detail in the grid, we examine two medium-sized establishments in apparel retail: one in India (establishment 1) and the other in Vietnam (establishment 2). Figure 3 plots the $MAX_{f,j}$ index in each business function for both establishments. For instance, establishment 1 uses a dynamic pricing system that automatically adjusts prices based on demand conditions, while establishment 2 uses an automated markup technology that collects information on costs and applies a uniform markup. The pricing business function includes five technologies (Figure A.7). From least to most sophisticated, these are: manual pricing (prices set without a formal account of the costs), automated markup, automated promotional pricing (prices adjusted based on seasonal factors), dynamic pricing, and personalized pricing (prices adjusted at the individual customer level using data analytics such as data mining and machine learning). The value of the MAX index for establishment 1 (4) indicates that it is one notch below the technology frontier in pricing, while the value for establishment 2 (2) shows that it is three notches below the frontier.

Establishment 1 also has a higher MAX level in other business functions such as merchandising and inventory. In these functions, it uses digital merchandising systems (DMS) and automated inventory controls, respectively. In contrast, establishment 2 selects products to display on shelves manually and uses a warehouse management system with specialized software.[19]

Nevertheless, establishment 2 has a higher MAX index than establishment 1 in other functions such as customer service, quality control, and sales. Specifically, establishment 2 attends to customer requests made online, checks product quality using statistical process control with software monitoring, and sells its products online using an external digital platform. In contrast, establishment 1 attends to customer requests over the phone, checks product quality manually with the support of digital technologies, and sells products directly at the establishment.

The variation in the relative technology sophistication rankings of establishments 1 and 2 across different business functions underscores the importance of a comprehensive coverage when characterizing the technological sophistication of an establishment. Focusing on one or a few functions or technologies provides an imprecise, and possibly biased, characterization of the sophistication of the technologies used in an establishment.

Next, we move from the establishment to the sector level and explore the cross-establishment distribution of technology sophistication in the food processing sector. We focus on the fabrication business function, which is relevant for all manufacturing establishments in FAT. Some of the technologies it covers, such as numerically controlled machines and robots, have been widely studied in automation research. The grid considers six classes of technologies. In increasing order of sophistication, these are (1) manual processes, (2) machines controlled by operators, (3) machines controlled by computers, (4) robots, (5) additive manufacturing including rapid prototyping and 3D printing, and (6) other advanced manufacturing processes such as laser, plasma sputtering, high-speed machine, E-beam and micro-machining).

The top right panel of Figure 4 plots the distribution of $MAX_{f,j}$ in fabrication across food processing establishments in South Korea. The histogram reveals a significant dispersion across establishments in the most sophisticated technologies available for production in fabrication. Establishments that process food using the world frontier's fabrication technologies coexist with others that just use manual processes. It is also worth noting that, in contrast to the popular perception, most establishments do not use robots or other more

[19] DMS is used to execute core merchandising activities, including product management, inventory replenishment, purchasing, vendor management, and financial tracking. One example of automated inventory controls is Computer Assisted Ordering (CAO), an inventory replenishment system that can use either sales or inventory algorithms to prepare a suggested reorder.

sophisticated technologies (i.e., $MAX_{f,j} \geq 3$) even in such an advanced economy as South Korea.[20]

To explore the cross-country differences in technology sophistication, the top panel of Figure 4 also plots the histogram of $MAX_{f,j}$ in Senegal (left) and India (middle). There are stark cross-country differences in the distribution. The mean and variance increase uniformly with the level of development in the country (i.e. Senegal, India, South Korea). Additionally, the cross-establishment distribution of $MAX_{f,j}$ is most skewed to the right in Senegal, and least in Korea.

The bottom panel in Figure 4 shifts the focus to the most widely used technology. In particular, it shows the histogram of $MOST_{f,j}$ in fabrication across food processing establishments for each country. By construction, the distribution of $MAX_{f,j}$ stochastically dominates the distribution of $MOST_{f,j}$, as $MAX_{f,j} \geq MOST_{f,j}$. However, the distributions of MAX and MOST differ significantly. For example, in 65% of Indian food processing establishments the most sophisticated technology used in fabrication is 'machines controlled by operators.' Yet, this technology is the most widely in only 35% of establishments.

The gap between MAX and MOST in this example motivates a deeper exploration of whether MAX and MOST are statistically distinct across a broad range of business functions and countries and, if that is the case, their relative importance in shaping the relationship between technology sophistication and productivity across establishments.

4 Technology Sophistication at the Business Function

We use the FAT dataset to examine technology at the business function level. We explore two issues: the range of technologies used in each function and the comparison between the most widely used and the most sophisticated technology available in the business function. We then compare our findings with the predictions of state-of-the-art models of technology in production and draw conclusions about relevant extensions.

4.1 The Array of Technologies Used in the Business Function

We begin our analysis of the vector of technologies used in a business function by counting the number of technologies from the grid that an establishment uses, denoted as $ANUM_{f,j}$. Table C.2 reports, for each function, the average $ANUM_{f,j}$ across all establishments that conduct the function in-house. On average, establishments use two different technologies

[20]This is also true in other sectors with greater penetration of robots such as the automotive sector. Even if we weight establishments by size (e.g., employment or sales) only 56% of establishments in automotive fabrication in South Korea use robots or a more sophisticated technology in fabrication.

per function. This average is consistent across both general and sector-specific business functions. The distribution of $ANUM_{f,j}$ reveals that 62.6% of functions use more than one technology, and 28.3% use at least three.

The vector of technologies used in a business function is relevant to two economic literatures. Schumpeterian models predict that when adopting a more sophisticated technology, an establishment will abandon the less sophisticated ones it was using. In models of technological leapfrogging, late adopters skip the less sophisticated technologies to directly use more sophisticated ones. Although FAT only provides cross-sectional data, it can be informative about the empirical support for these predictions. We explore the frequency of instances where establishments (i) completely skip or abandon less sophisticated technologies, (ii) use the least sophisticated technology despite having more advanced options, and (iii) create sophistication gaps by skipping some technologies in the business function.

Of the 37.4% of functions where only one technology is used, 52.8% use the least sophisticated technology. In the remaining 47.2%, the technology used is not the least sophisticated. Therefore, only in 18% (37.4% * 47.2%) of functions, establishments have fully skipped or abandoned all less sophisticated technologies. Conversely, in 70.4% of functions where multiple technologies are used, one of the technologies the establishments use is the least sophisticated technology in the grid. Finally, sophistication gaps are infrequent. Overall, they occur in 25% of business functions: 27% among GBFs and 17% among SSBFs. The GBFs where gaps are more frequent are payments (48%), business administration (34%), and sales (28%).

These observations show that establishments continue to use less sophisticated technologies even in business functions where more advanced technologies are available. FAT, therefore, does not support the predictions of Schumpeterian and leapfrogging models regarding the abandonment and skipping of less sophisticated technologies.

Additionally, the infrequency of sophistication gaps, combined with the regular use of the least sophisticated technology, implies that we can approximate the entire vector of technologies used in a business function just by the most sophisticated technology it uses. In other words, $MAX_{f,j}$ approximately captures the entire adoption history of the establishment in a business function.

To further explore the role of MAX in the technology upgrading process, we estimate the following regression:

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

where α_j and α_f are establishment and business function fixed effects. The point estimate of β , presented in column 2 of Table 4, is 0.84. The close to one-to-one movement of MAX and NUM suggests that the technologies that are introduced by an establishment in a business function are typically more sophisticated than existing ones.

4.2 The Intensity of Use of Technology

Technology research has mainly focused on the presence of new or advanced technologies in establishments, often neglecting the intensity of use of existing technologies. This focus is based on the belief that the technologies establishments use most intensively are the most sophisticated technologies they have adopted. Therefore, measuring the most widely used technology seems redundant, as $MAX_{f,j}$ is considered a sufficient statistic for $MOST_{f,j}$.

Departing from this tradition, we study the sophistication of the most widely used technology in business functions, $MOST_{f,j}$. Our primary objective is to determine whether, $MOST_{f,j}$ is indeed redundant, as suggested by the literature, or if it provides distinct insights into the sophistication of technologies used in a business function.

To explore this, we examine the MAX-MOST gap at the business function level. First, we document the magnitude and frequency of this gap. Second, we analyze the relationship between MOST and NUM, comparing it to the relationship between MAX and NUM. Third, we examine potential drivers of the MAX-MOST gap by studying its association with relevant proxies across establishments.

The MAX-MOST Gap. The average difference between $MAX_{f,j}$ and $MOST_{f,j}$ across establishments and functions is 0.68. This gap is significant economically and statistically given that sophistication measures range from 1 to 5, with standard deviations of 1.23 for $MAX_{f,j}$ and 1.09 for $MOST_{f,j}$. To study the frequency of MAX-MOST gaps, we calculate, for each establishment, the fraction of functions with multiple technologies that exhibit a MAX-MOST gap. The average frequency of the MAX-MOST gap across establishments is 0.62. The average is similar for GBFs (0.62) and SSBFs (0.61). However, the average frequency of MAX-MOST gaps in SSBFs varies significantly across sectors, ranging from 28% in health services to 82% in financial services. Except for health services and pharmaceuticals, all SSBFs have an average gap frequency of at least 50% (see Figure 5).

MAX-MOST gaps are also the norm across all countries, although there are significant differences, with average frequencies ranging from 51% in South Korea to 83% in Burkina Faso. Figure 6 shows that the average frequency of MAX-MOST gaps decreases with income, exhibiting a correlation of -0.55.

To further understand the differences between MAX and MOST, we regress $MOST_{f,j}$ on $MAX_{f,j}$ and study the fraction of the (within-establishment) variance in $MOST_{f,j}$ accounted for by $MAX_{f,j}$. Specifically, we estimate:

$$MOST_{f,j} = \alpha_j + \alpha_f + \beta * MAX_{f,j} + u_{f,j}, \tag{5}$$

where α_j and α_f are establishment and function effects.The estimates, reported in column 1 of Table 4, reveal that while $MAX_{f,j}$ and $MOST_{f,j}$ are positively correlated within establishments, $MAX_{f,j}$ explains only 34% of the variance in $MOST_{f,j}$. This indicates that $MAX_{f,j}$ is not a sufficient statistic for $MOST_{f,j}$.

To explore why this is the case, we regress $MOST_{f,j}$ on $NUM_{f,j}$ and compare this estimate with that of regressing $MAX_{f,j}$ on $NUM_{f,j}$ in specification (4). Specifically, we estimate

$$MOST_{f,j} = \alpha_j + \alpha_f + \beta * NUM_{f,j} + u_{f,j}. \tag{6}$$

The estimates, reported in column 3 of Table 4, imply that a 1-unit increase in $NUM_{f,j}$ is associated with an increase in $MOST_{f,j}$ by just 0.25, suggesting minimal impact of new technology adoption on the most widely used technology. This contrasts with the 0.84 estimate for $MAX_{f,j}$, showing that, the extension of the use of existing technologies in a business function and the adoption of new technologies are distinct technology upgrading processes and they are driven by different forces.

Drivers of the MAX-MOST Gap. A first step towards understanding the nature of the MAX-MOST gaps is to study whether they are transitory or permanent. Sluggishness in the extension of the use of new technologies could cause a transitory MAX-MOST gap.[21] Alternatively, the gap could reflect persistent factors that induce establishments to underutilize the most sophisticated technologies available in a function.

We consider the subsample of functions where establishments have adopted top-tier technologies, marked in bold in the grids presented in subsection A.1. For these technologies, FAT collects information on the year of adoption in the business function. We then divide this sample into two groups: those where $MAX_{f,j} = MOST_{f,j}$ and those where $MAX_{f,j} > MOST_{f,j}$. For each group, we examine the distribution of the years since adopting the top-tier technology in the function.Figure 7 shows that the distributions are similar, indicating that time is not a significant factor in closing the MAX-MOST gap, and that this gap is persistent.

We continue exploring the nature of the MAX-MOST gaps by considering three potential

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

drivers. First, establishments may struggle to extend the use of sophisticated technologies due to a lack of necessary worker skills, making human capital a limiting factor. Second, sophisticated technologies are often embodied in physical capital, and restricted access to capital may create a gap. Third, the gap may result from managerial mistakes rather than constraints. Worse managers or managers with imprecise or biased perceptions of their establishment's technological sophistication may tend to underutilize available technologies.

We define GAP_j as the number of functions with a MAX-MOST gap relative to the number of functions where the establishment uses multiple technologies. The establishment's human capital is measured by the fraction of college-educated employees, H_j. Restricted access to credit is measured by a dummy variable that reflects whether the establishment has been denied a loan application within the last year, $NOLoan_j$. We study the role of managerial mistakes with two different variables. The first is the management z-score that reflects the quality of management practices. The second is the manager's bias in technology perception, $Bias_j$, measured by the difference between the manager's assessment of the technology sophistication in the establishment and the actual technology sophistication.[22] We explore the relevance of these factors for the MAX-MOST gap by estimating the following specification:

$$GAP_j = \alpha_s+\alpha_c+\beta_0*Multiple_j+\beta_a*D_j^a+\beta_h*H_j+\beta_l*NOLoan_j+\beta_m*z-score_j+\beta_b*Bias_j+u_j, \quad (7)$$

where α_s are two-digit sector fixed effects, α_c are country fixed effects, $Multiple_j$ is the fraction of BFs where the establishment uses multiple technologies, and D_j^a are a set of dummy variables that capture the establishment age.

The estimates, reported in Table 5, support all the proposed drivers of the MAX-MOST gap. Having a loan rejected, a lower fraction of college educated workers, and a positive bias in the perceived sophistication of technology and worse management practices are all positively associated with the fraction of functions with a MAX-MOST gap in the establishment. Additionally, an establishment's age of at least 11 years is associated with a lower frequency of the gap. These estimates are robust to controlling by the fraction of functions where multiple technologies are used by the establishment.

4.3 Taking Stock

The way technology is conceptualized in production is crucial for many economic models, as it influences their predictions about optimal technology choices and the relationship between

[22] The specific question in FAT asks managers to rate the technology in their establishment relative to other establishments in the world. We convert this score to a 1-5 scale and compare it to $\overline{S}_j$ which is the simple average of MAX_j and $MOST_j$.

technology and productivity in establishments. The two main paradigms for this are Romer's (1990) love-for-variety model and Aghion and Howitt's (1992) quality ladder model. The findings in this section provide an empirical benchmark for evaluating these models and offer guidance on necessary extensions to more accurately depict technology sophistication at the business function level.

The fact that establishments typically use multiple technologies per business function and do not abandon less sophisticated technologies indicates that the technologies in the grid are not perfect substitutes.

The persistent gap between MAX and MOST suggests that establishments face significant costs in extending the use of sophisticated technologies already in place, beyond the costs of adopting new technologies. Our evidence indicates that these costs may be influenced by the establishment's access to skilled workers and capital. Additionally, managerial errors, reflected in worse management practices and in the bias in the manager's assessment of his establishment's sophistication, contribute to this gap. These factors create a wedge between MAX and MOST, as shown by the small fraction of MOST's variance explained by MAX and their differing associations with NUM. Notably, the MAX-MOST gap is larger in lower-income countries.

Existing frameworks of technology in production ignore the variation in the intensity of use as an important determinant of the technology sophistication of the business function. This omission limits their capacity to describe technology sophistication in the establishment and, based on the evidence provided in Section 6, their capacity to study the cross-establishment relationship between technology and productivity.

5 Technology Sophistication Across Establishments

We move from the business function to the establishment level to study the sophistication of technology across establishments. We are interested in two issues: the variation in technology sophistication across establishments and the association between technology sophistication and establishment characteristics.

Technology Sophistication in the Establishment. We measure technology sophistication in an establishment by averaging the technology sophistication across the business functions conducted in the establishment. This measure omits the sophistication of technologies used in functions that the establishment outsources to other establishments. This omission is not important if establishments outsource a small number of functions. As discussed in section 2, 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. Similarly, 87% of the relevant sector-specific business functions are conducted in-house. Therefore, the technology sophistication of in-house functions is a good proxy for the sophistication of the technologies establishments have access to both directly and indirectly via the sourcing of functions. Reassuringly, all the establishment-level findings we present next are robust to controlling for the fraction of functions an establishment conducts in-house (Cirera, Comin and Cruz, 2024).

Cross-establishment Variance in Technology Sophistication. Table 6 reports the key statistics of the cross-establishment distribution of technology sophistication. There is a large variation in technology sophistication across establishments. The standard deviation of MAX_j is 0.76 and the difference between the sophistication of the establishments in the 80^{th} and 20^{th} percentile of the distribution (p80-p20 gap) is 1.28. The standard deviation of $MOST_j$ is 0.63, and the p80-p20 gap is 1.16.

Technology sophistication varies significantly across countries. The difference between the average sophistication in the countries with highest and lowest levels are 1.53 for MAX and 1.01 for MOST. Figure 8 studies the relationship between technology sophistication and per capita income across countries. There is a strong positive correlation between per capita income and both measures of technology sophistication. For MAX the correlation is 0.78 and for MOST it is 0.94.

Technology sophistication also varies significantly within sectors. The standard deviation of MAX_j within sectors ranges from 0.89 in agriculture to 0.69 in manufacturing, while for $MOST_j$ it ranges from 0.68 in agriculture to 0.62 in manufacturing. Note that for both MAX and MOST, the sector with largest cross-establishment dispersion in technology sophistication is agriculture and the sector with smallest is manufacturing.

Technology sophistication also varies significantly within countries. For example, measuring the within-country dispersion in technology sophistication by the difference in sophistication between the establishments in the 80^{th} and 20^{th} deciles in a country, we find that the average p80-p20 gap across countries is 2.17 for MAX and 1.56 for MOST. However, this gap varies considerably across countries. Figure 9 shows the relationship between within-country dispersion in technology sophistication and per-capita income. The p80-p20 gap in a country is positively associated with per capita income. However, the strength of the association differs significantly between MAX and MOST. While the correlation of income with the p80-p20 gap of MAX is 0.33, for MOST it is 0.95.[23]

[23] If we measure the within-country dispersion in technology sophistication by the standard deviation, the correlations with (log) per capita income are 0.15 for MAX and 0.7 for MOST.

Cross-establishment Correlates of Technology Sophistication. We explore the establishments' characteristics that are associated with technology sophistication by estimating the following specification:

$$S_j = \alpha_c + \alpha_s + \beta * X_j + u_j \tag{8}$$

where $S_j = \{MAX_j, MOST_j\}$, α_c and α_s denote country and 2-digit sector fixed effects, respectively, and X_j reflects the characteristics of the establishment including fraction of employees with college degree, quality of management practices, size, age, exporter, multinational, and multi-establishment status.

Table 7 reports the estimates for MAX_j (column 1) and $MOST_j$ (column 2). We find that both measures of technology sophistication are positively associated with employees' human capital, the quality of management practices, larger establishment size, exporter, multinational, and multi-establishment status, and they have an inverted U-shape relationship with establishment age.

6 Technology Sophistication and Productivity

The relationship between technology and productivity is central to several important literatures. It is crucial to study the drivers of the large differences in productivity we observe across establishments and countries (e.g., Klenow and Rodríguez-Clare, 1997; Bartelsman, Haltiwanger and Scarpetta, 2013; Syverson, 2011). These cross-country differences in productivity are even more pronounced among agricultural establishments, as highlighted by the literature on the agricultural productivity gap (Caselli, 2005).

A natural hypothesis is that the cross-establishment variation in productivity reflects differences in technology sophistication across establishments. This answer prompts the question of why establishments implement technologies that differ so much in sophistication. The literature on appropriate technology suggests that this may be the case because establishments in low-income countries do not benefit from using sophisticated technologies as much as those in high-income economies. One reason for the cross-country heterogeneity in the marginal product of technology sophistication is that more advanced technologies may require complementary inputs that are relatively scarce in low-income countries. Note in any case that technology inappropriateness cannot explain the enormous variation in technology sophistication we observe within countries.

In this section, we use standard productivity regressions to explore the relationship between technology sophistication and productivity across establishments, thereby shedding light on these literatures.

6.1 Productivity Regressions

To explore the relationship between productivity and technology sophistication, we estimate variations of the following productivity regression:

$$Y_j = \alpha_{c,s} + \beta_k * K_j + \beta_h * H_j + \gamma * S_j + \theta * X_j + u_j \tag{9}$$

where the dependent variable is the log of sales per worker in establishment j, 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 measures of technology sophistication in the establishment, X_j is a vector of controls, $\alpha_{c,s}$ reflects various combinations of 2-digit sector and country dummies (typically not interacted), and u_j is classical measurement error.[24]

Columns 1 to 3 of Table 8 report the estimates of specification (9) with technology sophistication measured by the average of MAX_j and $MOST_j$, denoted by $\overline{S}_j$, and with both country and sector dummies (column 1), only sector dummies (column 2) and country-specific sector dummies (column 3). In all three specifications we find a large and positive coefficient of $\overline{S}_j$ in the productivity regression. In the baseline, with country and sector effects, an increase by one point in technology sophistication is associated with an increase in the (log) productivity of the establishment by roughly 0.5. The point estimate is roughly the same when allowing for country-specific 2-digit sector effect. However, the estimated coefficient of technology sophistication increases to 0.63 when excluding the country-fixed effects, suggesting that differences in technology sophistication are even more relevant to account for cross-country than within-country differences in productivity.

Since the establishment's productivity is measured by its sales per worker, one may wonder whether the association between technology sophistication and productivity is driven by the association with establishments' prices or with establishments' output per worker. FAT does not contain information on the prices of the goods and services produced by each establishment. However, it collects information on the markup charged for the main good or service sold by the establishment.[25] We ascertain the relevance of markups in the relationship between technology sophistication and productivity by including the markup as a control

[24]The estimates are robust to measuring productivity as value added per worker, using the log of sales as dependent variable including the log of employment as a control, or to calibrating the coefficients of capital and labor to the average sectoral share of the compensation to employment and capital in total sales (e.g., De Loecker and Syverson (2021)).

[25]This information is collected for Croatia, Chile, Brazil, Georgia, Vietnam, India, Cambodia, Bangladesh, Senegal and Ethiopia. As a result, controlling for markups reduces the sample from 13046 to 8553 establishments. The point estimate of the coefficient of $\overline{S}_j$ in the baseline specification for the subsample of establishments with markup information is 0.52, very similar to the point estimate in column 1.

variable. Column 5 of Table 8 reports the estimates. The markup in the main product is positively associated with sales per worker. However, controlling for the markup does not change the point estimate or the significance of the coefficient of $\overline{S}_j$ in the productivity regression. This finding supports the conclusion that the relationship between technology sophistication and productivity across establishment operates through output per worker rather than through prices.

Recent studies in productivity have highlighted the role of managerial practices (See Bloom and Van Reenen (2007) and Bloom et al. (2019). We explore the role of managerial practices in the relationship between technology sophistication and productivity by including the management practices z-score as a control. Consistent with Bloom et al. (2013), we estimate a positive coefficient for management practices (column 5). Its magnitude is relatively modest, as a one standard deviation increase in the management practices score is associated with an increase in establishment productivity by 6.2 percentage points. We further explore the possibility advanced by Bloom, Sadun and Reenen (2012) that technology sophistication and management practices are complementary. To this end, we introduce in the specification an interaction between $\overline{S}_j$ and a dummy that takes the value of 1 if the management score is above the median (column 6). We find that the coefficient of this interaction variable is positive and significant, suggesting that a key role of managers is the proper implementation of more sophisticated technologies. Importantly, controlling for the quality of management practices has little bearing on the estimated coefficient of technology sophistication, demonstrating the robustness of the relationship between technology sophistication and establishment productivity.

Linearity. The productivity regressions can help assess the appropriateness of the cardinalization used to construct the technology sophistication measures from the ordinal information extracted from FAT. An appropriate cardinalization of an ordinal variable is one that accurately captures its projection into a relevant cardinal variable. For technology sophistication, the most relevant variable is the (log) of productivity at the establishment level. Therefore, we can assess the appropriateness of the linear cardinalization used to construct $MAX_{f,j}$ and $MOST_{f,j}$ by exploring whether the relationship between (log) productivity and the measures of technology sophistication across establishments is approximately linear.

Table 9 explores the linearity of the relation between $\overline{S}_j$ and productivity. Column 1 reports the estimates of the productivity regression, allowing the coefficient of $\overline{S}_j$ to differ between establishments ranked above or below the median sophistication level. We find that the coefficient of the interaction between $\overline{S}_j$ and the "above median sophistication" dummy is negative but it is quantitatively small, and significant only at the 10% level.

Column 2 introduces greater flexibility in the specification by replacing the $\overline{S}_j$ with three dummies that reflect whether the establishment' sophistication falls in one of three intervals: [1.5-2.5), [2.5,3.5), and [3.5,5], leaving out the interval [1-1.5). These intervals are constructed so that they span the entire range of $\overline{S}_j$, and each contains a significant portion of the establishments in the sample. The estimated coefficients of these dummies imply that the increments in (log) productivity associated with a unitary increase in average sophistication are .42 (with an standard error of 0.05) when moving from the first to the second interval, .5 (s.e. 0.06) when moving from the second to the third, and .39 (s.e. 0.08) when moving from the third to the fourth.[26] These estimates demonstrate that the slope of the relationship between $\overline{S}_j$ and (log) productivity is roughly constant, and therefore, well approximated by a linear relationship.

This finding reassures us that the linear cardinalization used to construct the technology sophistication measures accurately represents the mapping from ordinal technology sophistication measures to establishment productivity.

Dimensions of technology sophistication. The variable $\overline{S}_j$ aggregates different dimensions of technology sophistication in the establishment. It includes both the sophistication of technologies in general and in sector-specific functions, as well as the MAX_j and $MOST_j$ measures. Next, we unpack $\overline{S}_j$ to explore which dimensions of technology sophistication most significantly impact the relationship between $\overline{S}_j$ and productivity across establishments.

We start by decomposing $\overline{S}_j$ into the average sophistication across the GBFs ($\overline{S}_{GBF,j}$) and across the SSBFs ($\overline{S}_{SSBF,j}$). Columns 1 and 2 of Table 10 report the estimates that result from replacing $\overline{S}_j$ with $\overline{S}_{GBF,j}$ and $\overline{S}_{SSBF,j}$ in specification (9). To ensure consistency in the coverage, in this particular exercise, we restrict the sample to establishments in sectors where technology information is recorded for both GBFs and SSBFs. We find that establishment productivity is strongly and positively associated with technology sophistication in both GBFs and SSBFs, but the association is stronger for GBFs with the coefficient for ($\overline{S}_{GBF,j}$ being roughly three times larger than that for $\overline{S}_{SSBF,j}$.

Next, we separate $\overline{S}_j$ into MAX_j and $MOST_j$. In the specification with country and sector fixed effects (column 3), both coefficients are positive and significant, indicating that both adopting new technologies and expanding the use of existing sophisticated technologies are associated with higher productivity. However, the coefficient for $MOST_j$ is six times larger than the coefficient for MAX_j, suggesting that the expansion of the use of existing technologies is much more relevant for productivity than the adoption of new technologies to the business function. This asymmetry is even more pronounced across countries. In the

[26] The increments in average $\overline{S}_j$ for each consecutive pair of intervals are 0.742, 0.881, 0.924.

specification without country effects (column 4), the coefficient for $MOST_j$ increases to 0.74 while MAX_j becomes insignificant.

Interpretation. In section 4, we showed that MOST and MAX represent distinct dimensions of technology upgrading. The estimates in Table 10 indicate that these processes are differently associated with productivity. The stronger association of $MOST_j$ with productivity as compared to MAX_j has significant implications. Positively, it suggests the need for theoretical frameworks that better link technology sophistication with productivity, emphasizing $MOST_j$. Normatively, it highlights that current innovation and technology policies, which focus on increasing MAX_j, should have a broader scope and also aim at expanding $MOST_j$.

Additionally, the differential association between establishment productivity and MAX_j and $MOST_j$ can shed light on the interpretation of the regression coefficients. In this paper, we avoid drawing causal interpretations from the associations between variables. Specifically, the estimates of the productivity regressions in Table 8 are consistent both with a productivity enhancing effect of sophisticated technologies and with an effect of establishment productivity on the return to implementing more sophisticated technologies. However, under this second interpretation, productivity should have relatively symmetric effects on the returns to adopting new technologies and to extending the use of existing technologies. Therefore, the strong asymmetry in the coefficient estimates for MAX_j and $MOST_j$ in Table 10 is more consistent with a causal effect of the extension of the use of sophisticated technologies on establishment productivity.

6.2 Development Accounting

Next, we use the estimates from the productivity regressions to conduct development accounting exercises. Specifically, we compute how much of the variation in productivity and revenue-based total factor productivity (TFPR) across establishments can be accounted for by differences in technology sophistication. To calculate the contribution to productivity dispersion, we regress the (log) of sales per worker and technology sophistication ($\overline{S}_j$) on the country and sector dummies included in the relevant specification of (9). We then residualize these variables and calculate the gap between the 10^{th} and 90^{th} percentiles. The contribution of factor technology sophistication to cross-establishment differences in productivity results from multiplying the 10-90 gap in residualized $\overline{S}_j$ by its coefficient in the productivity regression and dividing by the 10-90 gap in residualized productivity.

To compute the contribution to $TFPR$ dispersion, we residualize (log) sales per worker and $\overline{S}_j$ by the appropriate country and sector dummies, as well as by K_j and H_j. After obtaining the 10-90 gaps in the residualized productivity and technology sophistication variables, we follow the same procedure as before to determine the contribution of technology sophistication to the dispersion in TFPR.

Table 11 reports the results from the development accounting exercises. The first three rows correspond to the productivity regressions reported in the first three columns of Table 8. In the baseline specification with country and sector fixed effects, differences in technology sophistication account for 23% of the differences in productivity and 24% of the differences in TFPR across establishments. Excluding country effects allows us to study these contributions both within and across countries. In this case, differences in productivity account for 26% of productivity differences and 31% of TFPR differences across establishments.

6.3 The Agricultural Productivity Gap

Cross-country differences in productivity are roughly twice as large in agriculture than in non-agricultural sectors (Caselli, 2005). The FAT dataset is consistent with this so-called agricultural productivity gap as the gap between the (log) productivity of establishments in the 90^{th} and 10^{th} deciles is 5.91 in agriculture, compared to 4 in services. This implies that the 90-to-10 productivity ratio is 6.75 times (i.e. $exp(1.91)$) larger in agriculture than in services.

To study the role of technology sophistication in the agricultural productivity gap, we re-estimate the productivity regression (9) separately for each one-digit sector. The estimates, reported in Table 12 show a strong positive association between technology sophistication and productivity across all three sectors. However, the coefficient of technology sophistication varies significantly across sectors, being largest in agriculture and smallest in services.

Rows 4-9 of Table 11 report the contribution of technology sophistication to the dispersion in productivity and TFPR across establishments in each sector. Technology sophistication accounts for 50% of the dispersion in TFPR in agriculture, 30% in manufacturing and 28% in services. Within countries, it accounts for 33% of TFPR dispersion in agriculture, 26% in manufacturing and 24% in services. We find similar contributions to the cross-establishment dispersion in productivity (see column 1 of Table 11).

Differences across sectors in the contribution of technology sophistication to cross-establishment dispersion in productivity have implications for the agricultural productivity gap. In agriculture, technology sophistication accounts for a 90-10 log-productivity gap that is 1.05 points larger than in services. This means that over half of the agricultural productivity gap (1.05 out of 1.91 log-points) can be accounted for by differences in technology sophistication across establishments.

6.4 Appropriate Technology

The appropriate technology hypothesis has conjectured that establishments in poor countries do not extensively use sophisticated technologies because the scarcity of human and physical capital limits the potential productivity gains that sophisticated technologies embody (e.g., Basu and Weil, 1998; Acemoglu and Zilibotti, 2001). To formalize this hypothesis, suppose that the productivity of an establishment is given by $Y_j = A_c e^{\overline{S}_j}$, while the cost of implementing technology with sophistication $\overline{S}_j$ is $C_j(\overline{S}_j) = \frac{C_j}{2} e^{2\overline{S}_j}$. In this formulation, the marginal product of technology sophistication, A_c, may vary across countries reflecting the relative abundance of productive factors that are complementary to more sophisticated technologies. The parameter that captures the marginal cost of implementing more sophisticated technologies, C_j, potentially varies across establishments. Establishment j in country c chooses to implement a sophistication level $\overline{S}_j = ln(A_c/C_j)$. Note that both A_c and C_j affect the sophistication of technologies implemented. However, the marginal product of technology sophistication only depends on A_c. This insight allows us to explore the inappropriateness of more sophisticated technologies by studying whether the estimate of the marginal product of technology sophistication is larger in high- than in low-income countries.

To study this prediction, we split the FAT sample between the high-income countries (South Korea, Poland, and Croatia) and the rest. Th e latter group includes countries classified by the World Bank as low- and middle-income, is referred to low-income for brevity. We examine whether the coefficient of technology sophistication in the productivity regressions differs between these two groups. Table 13 presents the results. Column 1 includes a dummy for high income countries interacted with $\overline{S}_j$. Columns 2 and 3 estimate separate productivity regression for the establishments in high- and low-income countries, allowing all the coefficient and the sector dummies to vary between the two subsamples. Column 4 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. 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. This suggests that A_c is not lower in low-income countries than in high-income countries.

A potential concern with this interpretation is that the lack of a differential association between productivity and technology sophistication in high- vs. low-income country establishments may be due to an omitted variable. This could be the case if the omitted variable is more strongly correlated with either technology sophistication or productivity in low-income economies. One possible such variable is access to finance. Omitting this variable in the productivity regression could result in a larger coefficient for technology sophistication in low-income economies than if it were properly controlled for. This bias could mask a ‘true’ lower value of A_c in low-income economies.

To explore whether the estimates in columns 1-4 reflect omitted variable bias or correctly reflect that A_c is not lower in low-income economies, we split the sample of establishments along potential proxies for the omitted variable. We then examine if there is a differential association between $\overline{S}_j$ and productivity within the subsamples for high- and low-income economies. If no differential association is observed, the case for omitted variable bias is weakened.

We consider two proxies for the omitted variables: the establishments' human capital and size, respectively measured by the fraction of college-educated workers and the number of employees. We split establishments between those above and below the median fraction of college-educated workers (columns 5-6 of Table 13) and those above and below the (unweighted) median number of employees (columns 7-8 of Table 13). In all four subsamples, we find that the coefficient of $\overline{S}_j$ is not higher in high-income economies than in low-income economies. This finding suggests that the failure to find a stronger association between technology sophistication and productivity in high-income countries is unlikely to be due to the omission of variables that affect productivity more in low-income economies. This leads us to conclude that our findings suggest that more sophisticated technologies are generally appropriate for use in all countries, regardless of their development level.

7 Conclusions

This paper presents a new approach to comprehensively characterize the technologies used in an establishment. Introduces a tool, the grid, that describes the key business functions involved in production and the possible technologies to perform the main tasks in each function. We have implemented this methodology and assembled a dataset covering over 21,000 establishments in 15 countries at all stages of development. An exploration of the FAT dataset has uncovered three main findings. First, the most widely used technology in a business function (MOST) typically is not the most sophisticated technology available (MAX). This gap between MAX and MOST is not transitory. It reflects the different nature and dynamics of the two upgrading processes: adoption versus extension of the use of an adopted technology. Second, there are large differences in technology sophistication across establishments. Factors associated with greater technology sophistication include the establishment size, the human capital of its employees, the quality of the managerial practices, being an exporter, and being part of a multinational or a multi-establishment firm. Third, there is a strong and robust cross-establishment association between technology sophistication and productivity both within and between countries. This relationship (i) is linear, (ii) is much stronger for MOST than for MAX, (iii) accounts for more than 30% of the differences in productivity between establishments, and (iv) for 50% between agricultural establishments, and (v) the association between establishment productivity and technology sophistication is not weaker in low- than in high-income countries.

We plan to build on the methodological and empirical contributions of this paper in various directions. First, we intend to extend FAT to collect data in more countries and also to create grids for tasks in new sectors that allow us to provide a detailed technological account of more establishments.

Second, the distinct nature of MAX and MOST documented in this paper deserves scrutiny. On the theoretical side, we plan to develop frameworks that connect productivity to the range of technologies used but also to how intensively they are used and that rationalize the observed gap between MAX and MOST. On the empirical front, it seems of first-order importance to assess the relevance of the possible explanations for the gap.

A separate issue that we have overlooked in this paper is the aggregation of technology sophistication across business functions into an establishment-level sophistication index. In this paper, we have taken the reasonable shortcut for a descriptive exercise of constructing establishment-level measures of technology sophistication as the simple average of the function-level sophistication measures. However, developing a theory that rationalizes the variation between the functions of an establishment in technology sophistication and that provides a foundation for the construction of measures of technology sophistication at the establishment level would be a major development.

In this paper, we have intentionally avoided studying the range of business functions conducted in establishments (i.e. the horizontal dimension of the grid). This topic, along with developing models that rationalize the relationship between technology sophistication and the limits of the establishment and its specialization from a task perspective, warrants separate attention.

References

Acemoglu, Daron, and David Autor. 2011. "Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings." In . Vol. 4 of *Handbook of Labor Economics*, , ed. David Card and Orley Ashenfelter, 1043 – 1171. Elsevier.

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

Acemoglu, Daron, and Pascual Restrepo. 2018. "The race between man and machine: Implications of technology for growth, factor shares, and employment." *American economic review*, 108(6): 1488–1542.

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.

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.

Battisti, Giuliana, and Paul Stoneman. 2003. "Inter-and intra-firm effects in the diffusion of new process technology." *Research Policy*, 32(9): 1641–1655.

Battisti, Giuliana, Heinz Hollenstein, Paul Stoneman, and Martin Woerter. 2007. "Inter And Intra Firm Diffusion Of Ict In The United Kingdom (Uk) And Switzerland (Ch) An Internationally Comparative Study Based On Firm-Level Data." *Economics of Innovation and New Technology*, 16(8): 669–687.

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, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does management matter? Evidence from India." *The Quarterly journal of economics*, 128(1): 1–51.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen. 2019. "What Drives Differences in Management Practices?" *American Economic Review*, 109(5): 1648–1683.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "Americans do IT better: US multinationals and the productivity miracle." *American Economic Review*, 102(1): 167–201.

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.

Boothby, Daniel, Anik Dufour, and Jianmin Tang. 2010. "Technology adoption, training and productivity performance." *Research Policy*, 39(5): 650–661. Special Section on Government as Entrepreneur.

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.

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

Cirera, Xavier, Diego A Comin, and Marcio Cruz. 2024. "Anatomy of Technology and Tasks in the Establishment." National Bureau of Economic Research Working Paper 32281.

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.*

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 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.

De Loecker, Jan, and Chad Syverson. 2021. "An industrial organization perspective on productivity." In *Handbook of industrial organization.* Vol. 4, 141–223. Elsevier.

Dunne, Timothy. 1994. "Plant Age and Technology use in U.S. Manufacturing Industries." *The RAND Journal of Economics*, 25(3): 488–499.

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, Gene M, and Esteban Rossi-Hansberg. 2008. "Trading tasks: A simple theory of offshoring." *American Economic Review*, 98(5): 1978–1997.

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.

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

Hollenstein, Heinz, and Martin Woerter. 2008. "Inter- and intra-firm diffusion of technology: The example of E-commerce: An analysis based on Swiss firm-level data." *Research Policy*, 37(3): 545–564.

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

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

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

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.

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.

Wood, Lexi. 2024. "How Much Does ERP Cost?" https://softwareconnect.com/learn/erp-pricing/.

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

Zeira, Joseph. 1998. "Workers, machines, and economic growth." *The Quarterly Journal of Economics*, 113(4): 1091–1117.

Figures and Tables

Figure 1: General Business Functions and Their Technologies

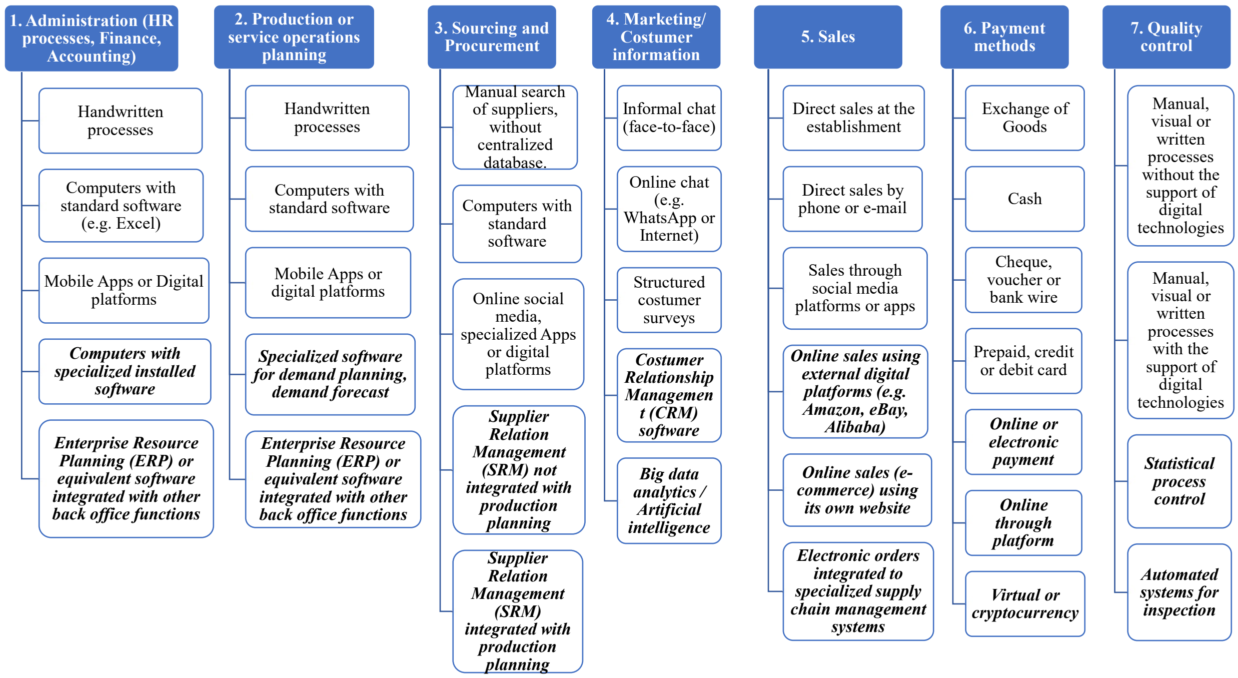


Figure 2: Sector Specific Business Functions and Technologies in Agriculture

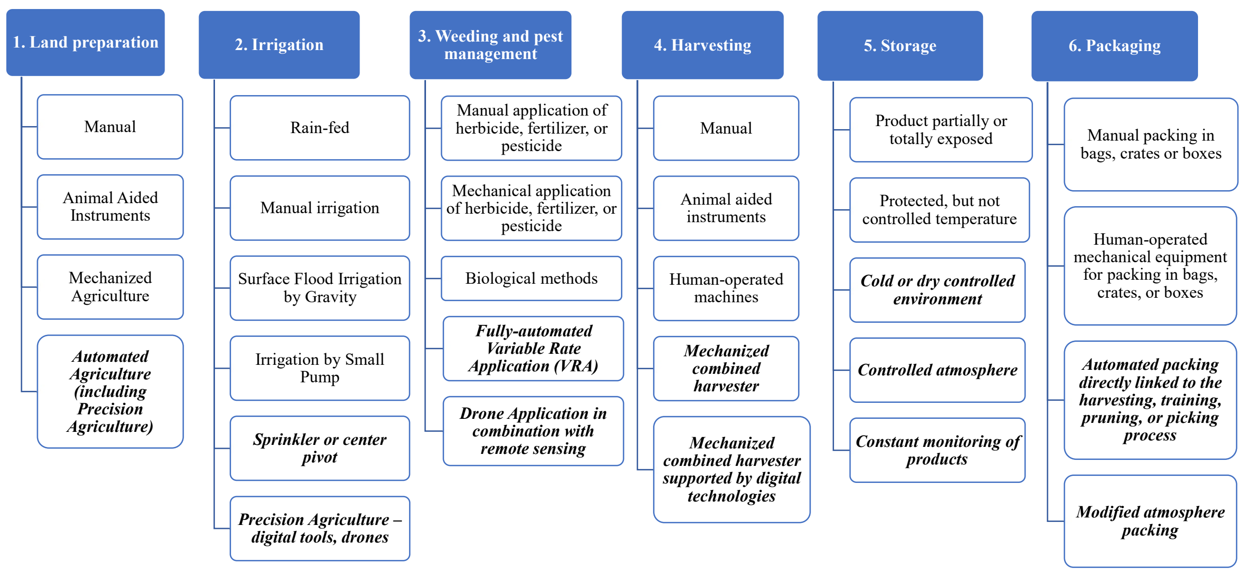
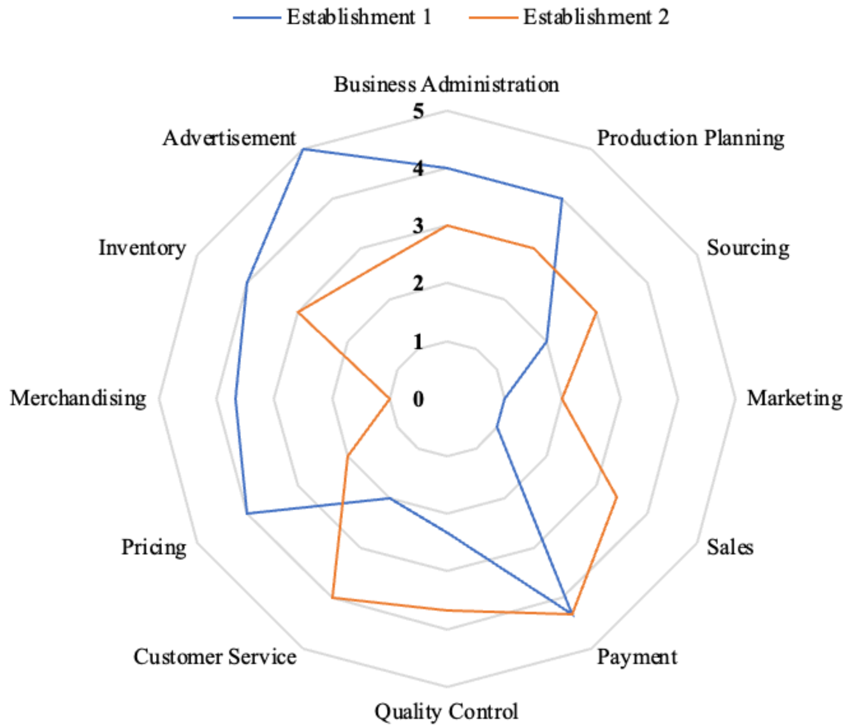
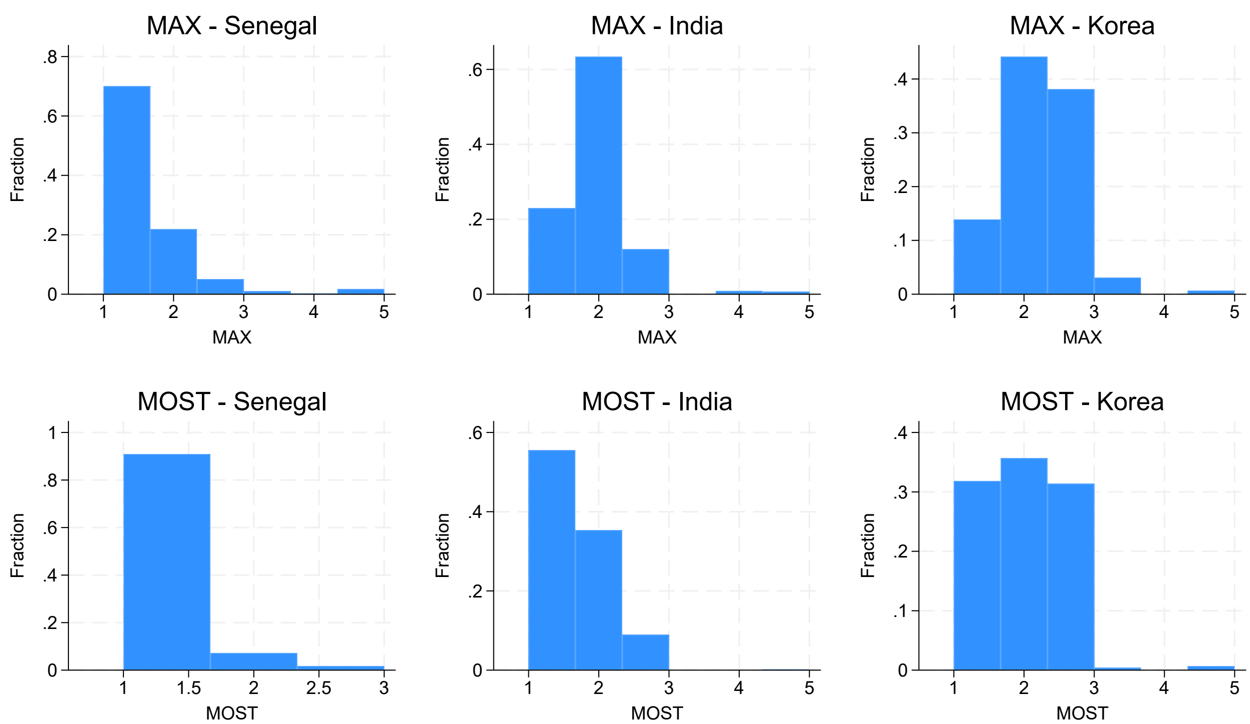


Figure 3: $MAX_{f,j}$ in two establishments in retail services



Note: Figure displays the technology index $MAX_{f,j}$ across all business functions for two individual establishments in retail services.

Figure 4: Distribution of technology sophistication in Food Processing (Fabrication)



Note: Figure displays the distribution of the technology measures $MAX_{f,j}$ and $MOST_{f,j}$ for the fabrication function across establishments in the food processing manufacturing sector in Senegal, India, and Korea. Each column in the histograms corresponds to one technology. From least to more sophisticated these are: (i) manual processes, (ii) machines controlled by operators, (iii) machines controlled by computers, (iv) robots, (v) additive manufacturing including rapid prototyping and 3D printers, and (vi) other advanced manufacturing processes such as plasma sputtering, high speed machine, E-beam, and micromachining.

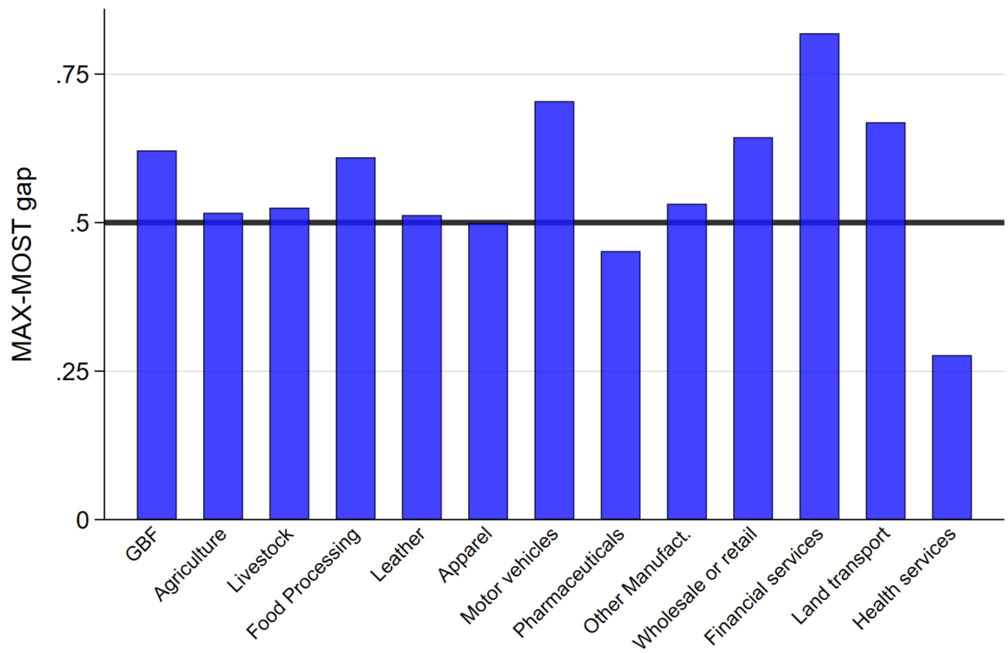
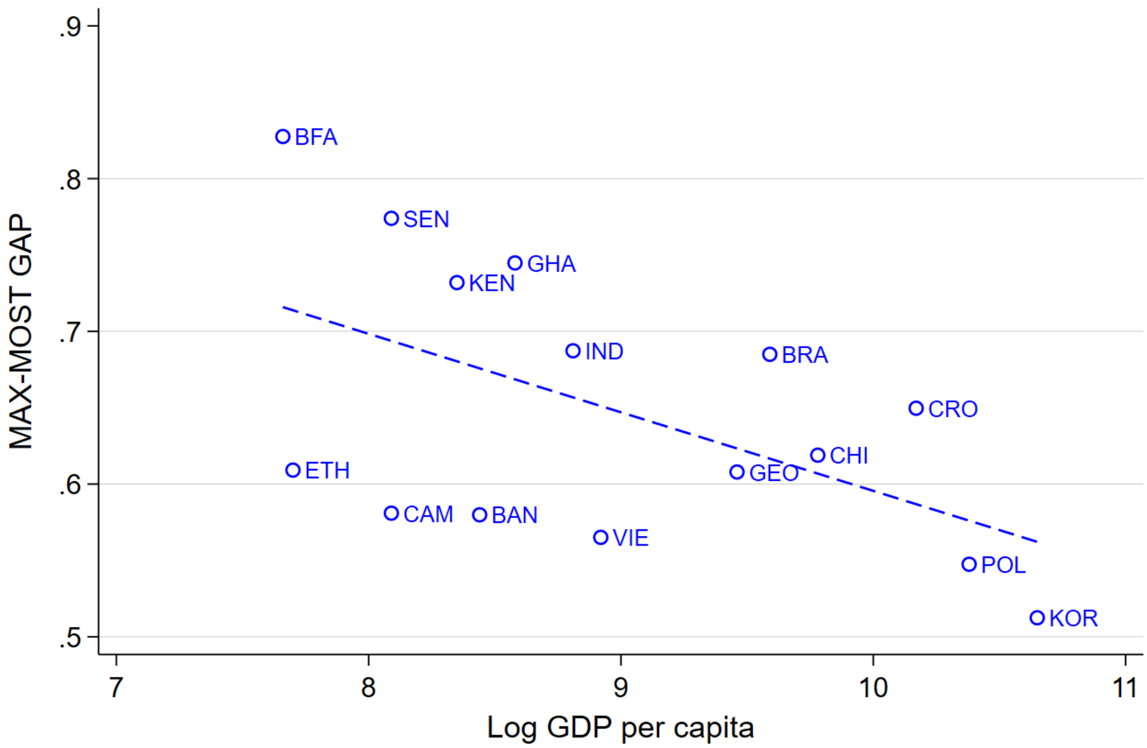


Figure 5: Average MAX and MOST GAP by Class of Business Function

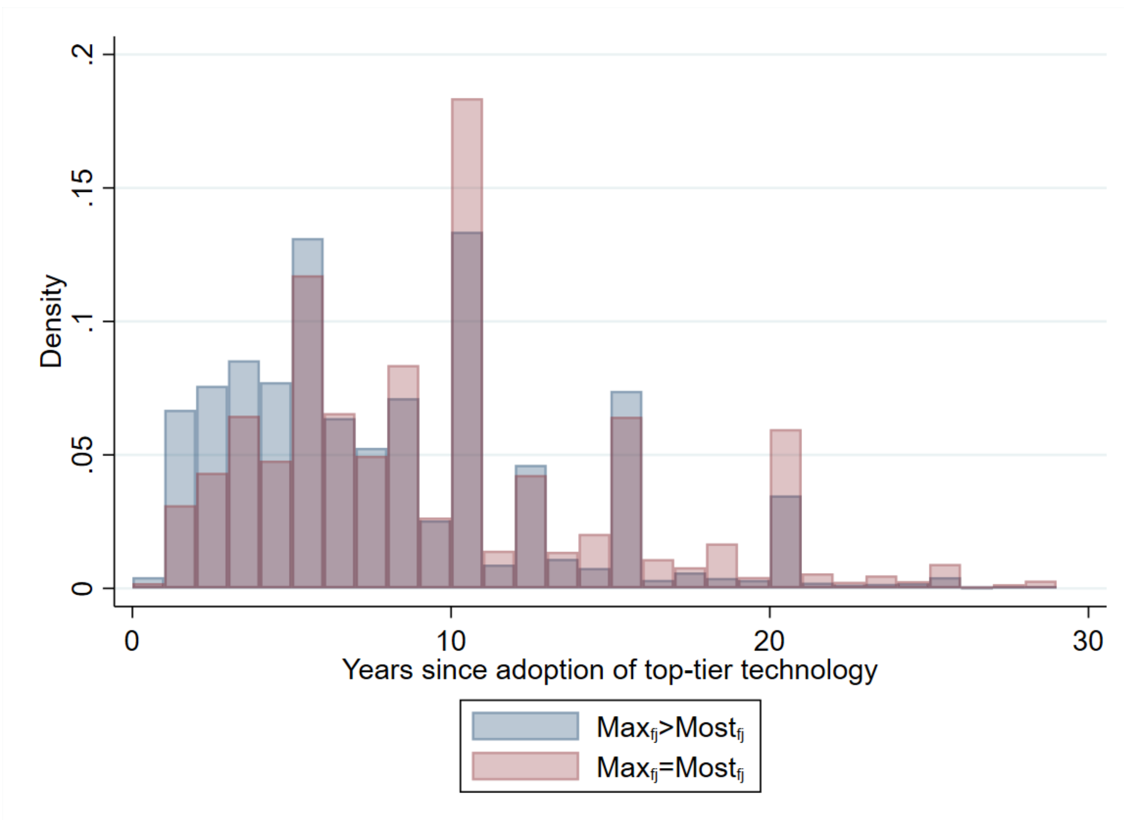
Note: Figures displays the average across relevant establishments of the share of business functions f in a given class of functions (e.g., GBF or SSBF in the sector) where $MAX_{f,j} > MOST_{f,j}$ and $NUM_{f,j} > 1$.

Figure 6: MAX-MOST GAP across countries



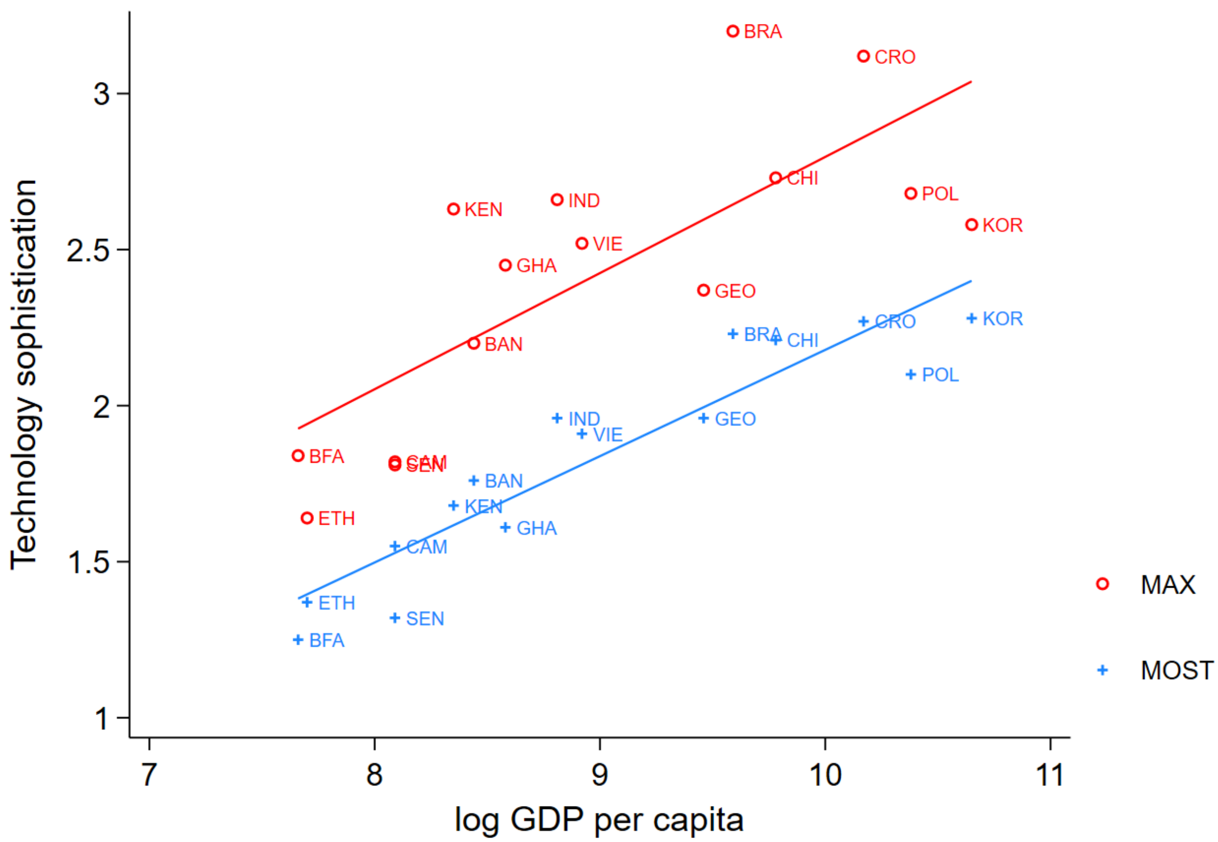
Note: The country-level MAX-MOST Gap is the average MAX-MOST Gap across the establishments in the country.

Figure 7: Distribution of Years since Adoption of Top-tier Technologies Conditional on MAX-MOST Gap



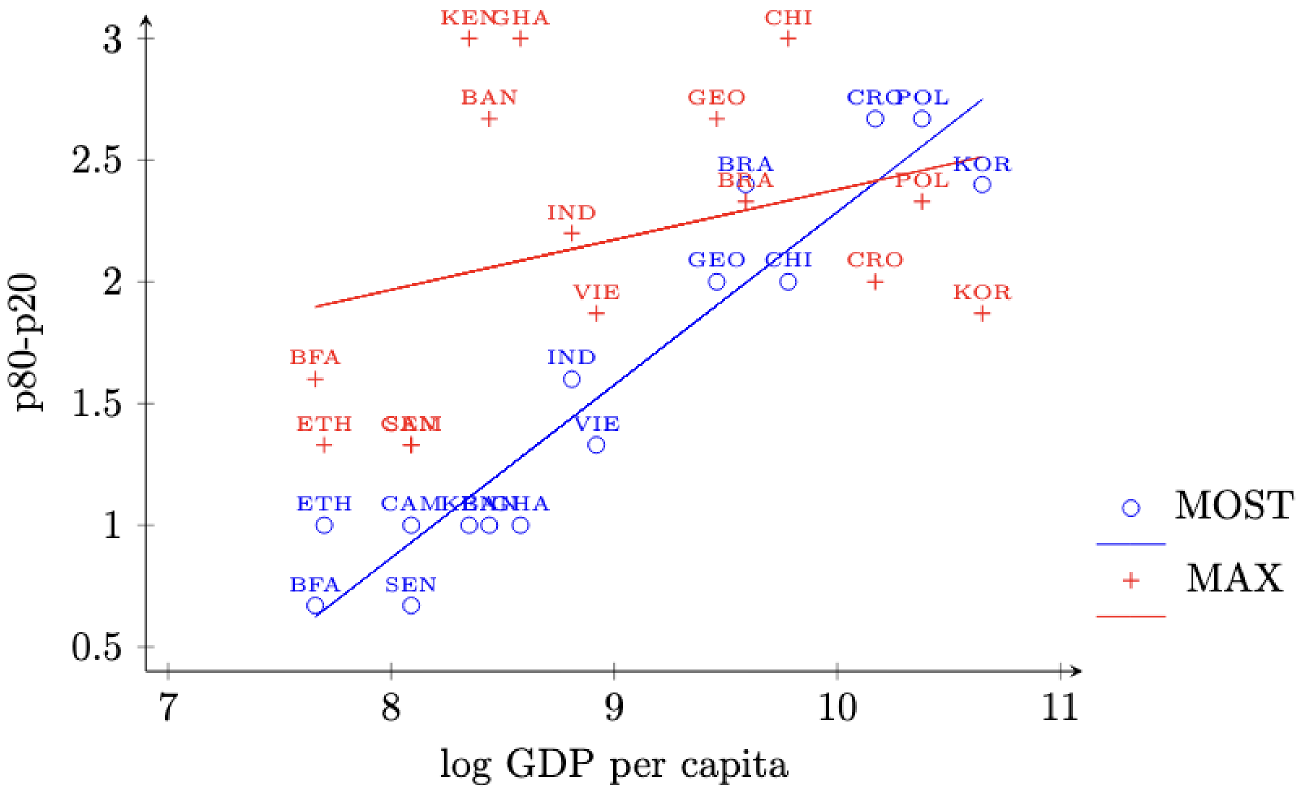
Note: Top-tier technologies are listed in Appendix A. Only functions where the establishment uses multiple technologies are considered. MAX-MOST gap is present in business function if $MAX_{fj} > MOST_{fj}$. It is absent otherwise.

Figure 8: MAX and MOST across countries



Note: Country-level MAX and MOST are, respectively, the weighted averages of establishment-level MAX_j and $MOST_j$, where the weights are sampling weights.

Figure 9: Within-country dispersion in technology sophistication and per capita income



Note: P80-p20 is the difference between the technology sophistication (MAX or MOST) of establishments in the 80^{th} and 20^{th} percentile of technology sophistication in the country. Percentiles are computed using establishment weights.

Table 1: Comparison of Technology Categories: Business Administration

	Handwritten process	**Standard Software**	**Mobile apps**	**Specialized Software**	**Enterprise Resource Planning (ERP)**
Functionality	Basic manual tasks (e.g., simple bookkeeping, and employee records).	Handles financial, accounting, and HR record, with manual inputs or built in functions.	Pre-designed to handle financial, accounting, and HR record. Limited scale and customization.	Extensive specialized tools for complex financial, accounting, and HR management.	Comprehensive management of all finance, accounting, and HR processes.
Integration	No.	Limited. It requires manual processing or additional templates.	Good integration capabilities with limited customization.	Integration with systems and customizable reporting tools.	Full integration with a wide range of functions and customization.
Automation	No.	Limited. It requires manual scripting.	Good automation for specific processes with limited scale.	High-level of automation within their specialized domains.	High-level of automation across all functions.
Experts ranking	1	2	3	4	5
Reason for Ranking	Manual processes with minimal functionality, no automation, and no integration capabilities.	Basic functionality and some level of automation. Limited integration capabilities.	Good functionality, integration, and automation, but limited scale and customization.	High functionality, integration, and automation capabilities within specialized domains.	Comprehensive functionality, full integration, and advanced automation.
Technology Example	Paper Ledger	Microsoft Excel, Google Sheets	QuickBooks Online, BambooHR	Oracle Financials, Intuit QuickBooks (Desktop), Workday HCM	SAP ERP, Oracle NetSuite
Cost for acquiring the technology	Negligible	Microsoft Excel: $159.99; Google Sheets (Free-$18/user/month)	QuickBooks Online: $30-$200/month; BambooHR: $108/month for 20 employees	Oracle Financials: $600+/month; Intuit QuickBooks: $1,481+/year/user;	SAP ERP; Oracle NetSuite price varies. Average ERP $1,740-$9,330/month.
Launch year	Pre- 1900	Microsoft Excel: 1985; Google Sheets: 2006	QuickBooks Online: 2001; BambooHR: 2008	Oracle Financials: 1989; Intuit QuickBooks: 1998; Workday HCM: 2006	SAP ERP: 1981; Oracle NetSuite: 1998
ChatGPT ranking	1	2	3	4	5
ChatGPT time (task)*	5 hours	2 hours	1.5 hours	1 hour	30 min

Sources: Product description on the websites of various companies, including Microsoft, Google, QuickBooks, Bamboo HR, Oracle, SAP, and Workday. Wood (2024) provides estimates of average costs for ERP software. The prompts for ChatGPT ranking and estimated time to perform a typical task – manage payroll and prepare financial statements – are available in the appendix.

Table 2: 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 3: Average level of technology measures

	$ANUM_{f,j}$	$NUM_{f,j}$	$MAX_{f,j}$	$MOST_{f,j}$	N_f
ABFs	2.1	2.1	2.7	2.0	4.8
GBFs	2.1	2.0	2.6	2.0	5.3
SSBFs	2.1	2.1	2.7	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 4: Relationship between technology measures

	$MOST_{f,j}$	$MAX_{f,j}$	$MOST_{f,j}$
$MAX_{f,j}$	0.55^{***}		
	(0.01)		
$NUM_{f,j}$		0.84^{***}	0.25^{***}
		(0.01)	(0.01)
N	187497	187497	187497
R-squared	0.66	0.75	0.50
BF FE	Y	Y	Y
Firm FE	Y	Y	Y
Variation Explained	0.34	0.47	0.05

Notes : This table reports the regression estimates of specifications 5, 6, and 4. 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 5: Cross-establishment Drivers of MAX-MOST Gap

	GAP_j	
	(1)	(2)
H_j	-0.07***	-0.06***
	(0.01)	(0.01)
$NOLoan$	0.04***	0.05***
	(0.01)	(0.01)
$Bias$	0.01***	0.01***
	(0.00)	(0.00)
Management (Z-Score)	-0.00***	-0.00***
	(0.00)	(0.00)
Age: 6 - 10 Years	-0.01	-0.00
	(0.01)	(0.01)
Age : 11 - 15 Years	-0.04***	-0.04***
	(0.01)	(0.01)
Age : 16+ years	-0.02***	-0.02***
	(0.01)	(0.01)
$Multiple$		-0.08***
		(0.01)
N	16605	16605
R-squared	0.10	0.11
2-Dig. Sector FE	Yes	Yes
Country FE	Yes	Yes

Notes : MAX-MOST Gap, GAP_j, 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. $NOLoan$ is a binary variable taking value 1 if the establishment has any loan rejected in the last year. $Bias$ is defined as $[(1 + 4/9 * SelfPerceptionScore) - S_j]$, where $SelfPerceptionScore$ asks the establishments their perceived technology ranking (scale from 1 to 10) as compared to the whole world. $Multiple$ is the fraction of BFs in the establishment with $NUM_{f,j} > 1$. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table 6: Cross-establishment distribution of MAX_j and $MOST_j$

	MAX_j					$MOST_j$				
Sector	Mean	SD	p20	p50	p80	Mean	SD	p20	p50	p80
Overall	2.61	0.76	1.99	2.52	3.27	2.02	0.63	1.47	1.93	2.53
Agriculture	2.63	0.89	1.81	2.51	3.73	2.07	0.68	1.37	2.03	2.68
Manufacturing	2.61	0.69	2.02	2.53	3.18	2.01	0.62	1.47	1.92	2.52
Services	2.61	0.79	1.97	2.51	3.30	2.02	0.64	1.47	1.93	2.52

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

Table 7: Technological sophistication and establishment characteristics

	(1) MAX_j	(2) $MOST_j$
H_j	0.40***	0.24***
	(0.02)	(0.01)
Management (Z-Score)	0.13***	0.11***
	(0.00)	(0.00)
Size: Medium	0.31***	0.22***
	(0.01)	(0.01)
Size: Large	0.66***	0.46***
	(0.02)	(0.02)
Age: 6 to 10	0.09***	0.13***
	(0.01)	(0.01)
Age: 11 to 15	0.03**	0.14***
	(0.01)	(0.01)
Age: 16+	0.00	0.03***
	(0.01)	(0.01)
Foreign owned	0.25***	0.25***
	(0.02)	(0.01)
Exporter	0.19***	0.15***
	(0.02)	(0.01)
Multi-establishment	0.27***	0.18***
	(0.01)	(0.01)
N	17161	17161
R-squared	0.46	0.38
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 8: Productivity and Technology Sophistication

	ln(Sales per worker)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
K_j	0.234***	0.266***	0.223***	0.300***	0.304***	0.233***	0.231***
	(0.007)	(0.007)	(0.007)	(0.009)	(0.009)	(0.007)	(0.007)
H_j	0.191***	0.585***	0.150***	0.084*	0.109**	0.189***	0.202***
	(0.040)	(0.042)	(0.040)	(0.048)	(0.049)	(0.040)	(0.040)
$\overline{S}_j$	0.493***	0.631***	0.504***	0.527***	0.530***	0.460***	0.422***
	(0.019)	(0.020)	(0.019)	(0.024)	(0.024)	(0.020)	(0.022)
Markup					0.189***		
					(0.052)		
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***	5.484***	5.208***	6.206***	6.217***
	(0.175)	(0.147)	(0.310)	(0.203)	(0.217)	(0.175)	(0.175)
N	13046	13046	13046	8553	8553	13046	13046
R-squared	0.407	0.234	0.435	0.341	0.342	0.409	0.410
Sector FE	Yes	Yes		Yes	Yes	Yes	Yes
Country FE	Yes	No		Yes	Yes	Yes	Yes
Sector x Country FE			Yes				

Notes : Estimates of specification (9). D(High Management) is a dummy that takes the value 1 if the management z-score of the establishment is above the median, and 0 otherwise. Markup is the gross markup (1+markup%)for the main product or service produced in this establishment. Columns (4) and (5) are calculated only for the sample where the markup data is collected. All regressions estimated using establishment-level sampling weights. *, ** and *** denote 10%, 5% and 1% significance respectively.

Table 9: Linearity of Relationship Between Productivity and Technology Sophistication

	ln(Sales per worker)	
	(1)	(2)
K_j	0.23***	0.24***
	(0.01)	(0.01)
H_j	0.19***	0.24***
	(0.04)	(0.04)
$\overline{S}_j$	0.56***	
	(0.04)	
$\overline{S}_j$ * D(High Sophistication)	-0.03*	
	(0.02)	
D($1.5 \leq \overline{S}_j \leq 2.5$)		0.31***
		(0.04)
D($2.5 \leq \overline{S}_j \leq 3.5$)		0.75***
		(0.05)
D($\overline{S}_j \geq 3.5$)		1.11***
		(0.07)
Constant	6.02***	6.77***
	(0.18)	(0.17)
N	13046	13046
R-squared	0.41	0.40
Sector FE	Yes	Yes
Country FE	Yes	Yes

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

Table 10: Productivity and Dimensions of Technology Sophistication

	ln(Sales per worker)			
	(1)	(2)	(3)	(4)
K_j	0.254***	0.287***	0.235***	0.267***
	(0.008)	(0.009)	(0.007)	(0.007)
H_j	0.119**	0.6482***	0.233***	0.633***
	(0.053)	(0.054)	(0.041)	(0.042)
$\overline{S}_{GBF,j}$	0.304***	0.452***		
	(0.028)	(0.029)		
$\overline{S}_{SSBF,j}$	0.081***	0.170***		
	(0.024)	(0.026)		
MAX_j			0.084***	-0.037
			(0.023)	(0.023)
$MOST_j$			0.433***	0.740***
			(0.025)	(0.027)
Constant	5.958***	5.594***	6.103***	5.971***
	(0.184)	(0.156)	(0.174)	(0.145)
N	8877	8877	13046	13046
R-squared	0.407	0.254	0.410	0.250
Sector FE	Yes	Yes	Yes	Yes
Country FE	Yes	No	Yes	No

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

Table 11: Development Accounting

Sector	Country FE	Contribution to	
		log of Sales per worker	TFPR
Overall	Y	0.23	0.24
	N	0.28	0.31
	Sector X Country	0.23	0.25
Agriculture	Y	0.30	0.33
	N	0.44	0.50
Manufacturing	Y	0.27	0.26
	N	0.33	0.30
Services	Y	0.20	0.24
	N	0.26	0.28

Notes : The table reports the contribution of $\overline{S}_j$ to the cross-establishment dispersion in productivity, (log) sales per worker, and TFPR, as discussed in the text. Cross-establishment dispersion is measured by gap between the establishments in 90^{th} and 10^{th} deciles of the distribution of relevant variable. The first three rows correspond to the estimates from the first three columns of Table 8. Rows 4 through 9 report contributions from sectoral regressions reported in Table 12.

Table 12: Productivity and Technology Sophistication - Across Sectors

	ln(Sales per worker)					
	(1)	(2)	(3)	(4)	(5)	(6)
K_j	0.342***	0.442***	0.234***	0.281***	0.218***	0.244***
	(0.023)	(0.024)	(0.008)	(0.009)	(0.011)	(0.012)
H_j	0.507**	0.825***	0.164***	0.604***	0.165***	0.586***
	(0.227)	(0.236)	(0.061)	(0.063)	(0.058)	(0.060)
$\overline{S}_j$	0.648***	1.023***	0.584***	0.701***	0.458***	0.587***
	(0.088)	(0.086)	(0.023)	(0.025)	(0.030)	(0.030)
Constant	5.402***	3.067***	6.471***	6.326***	7.502***	7.410***
	(0.419)	(0.233)	(0.124)	(0.108)	(0.420)	(0.139)
N	825	825	6032	6032	6189	6189
R-squared	0.716	0.577	0.480	0.327	0.382	0.186
2 Dig. Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	No	Yes	No	Yes	No
Data	Agri.	Agri.	Manu.	Manu.	Serv.	Serv.

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

Table 13: Technology Appropriateness

	ln(Sales per worker)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K_j	0.23***	0.11***	0.31***	0.24***	0.21***	0.26***	0.24***	0.23***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
H_j	0.20***	0.45***	0.09**	0.25***	0.54***	0.48*	0.67***	0.10*
	(0.04)	(0.09)	(0.05)	(0.04)	(0.07)	(0.29)	(0.07)	(0.05)
$\overline{S}_j$	0.51***	0.47***	0.50***		0.37***	0.63***	0.34***	0.54***
	(0.02)	(0.04)	(0.02)		(0.04)	(0.03)	(0.03)	(0.03)
D($1.5 \leq \overline{S}_j \leq 2.5$)				0.31***				
				(0.04)				
D($2.5 \leq \overline{S}_j \leq 3.5$)				0.78***				
				(0.05)				
D($\overline{S}_j \geq 3.5$)				1.21***				
				(0.08)				
$\overline{S}_j$ * D(High Income)	-0.07				0.08	-0.36***	-0.13 **	-0.00
	(0.04)				(0.07)	(0.07)	(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	6.08***	11.34***	5.31***	6.77***	6.56***	5.51***	6.55***	6.03***
	(0.18)	(0.36)	(0.18)	(0.17)	(0.41)	(0.20)	(0.22)	(0.31)
N	13046	2104	10942	13046	5803	6874	6383	6663
R-squared	0.41	0.30	0.38	0.40	0.39	0.47	0.43	0.42
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Data	All	High Income	Low Income	All	High H_j	Low H_j	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. 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. High H_j and low H_j are based on above and below median fraction of college-educated workers. 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

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 frames 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 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	305	63	Yes
Manufacturing Technology Survey (MTS)	17	0	No
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) 2018 Technology module	10	0	No
Annual Business Survey (ABS) 2019 Technology module	5	0	No

Note: The Number of technologies and business functions are computed by authors. MTS, ICTS, and ABS were conducted by the United States Census Bureau. SAT was conducted by the Statistics Canada. ICT Usage in Enterprise is conducted by EUROSTAT.

The FAT survey addresses important knowledge gaps compared to other surveys measuring technology at the firm or establishment level. To start, 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, financial statements and performance, which allow us to compute labor productivity and other measures at the establishment level.

A.1.1 The Grid

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.[27] 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.

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

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 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 in Section 2 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,accounting), 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.

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. Identifying key business functions and the frontier in each sector required 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 phases of data collection. These figures complement the information provided in Section 2, particularly Figure 2, which describes the functions and associated technologies for SSBFs in agriculture, among SSBFs. The complementary information is provided for all SSBFs, including Livestock (Figure A.1), Food Processing (Figure A.2), Wearing Apparel (Figure A.3), Leather and Footwear (Figure A.4), Automotive (Figure A.5), Pharmaceutical (Figure A.6), Wholesale and Retail (Figure A.7), Transportation (Figure A.8), Financial Services (Figure A.9), Health Services (Figure A.10), and Other Manufacturing (Figure A.11)). [28]

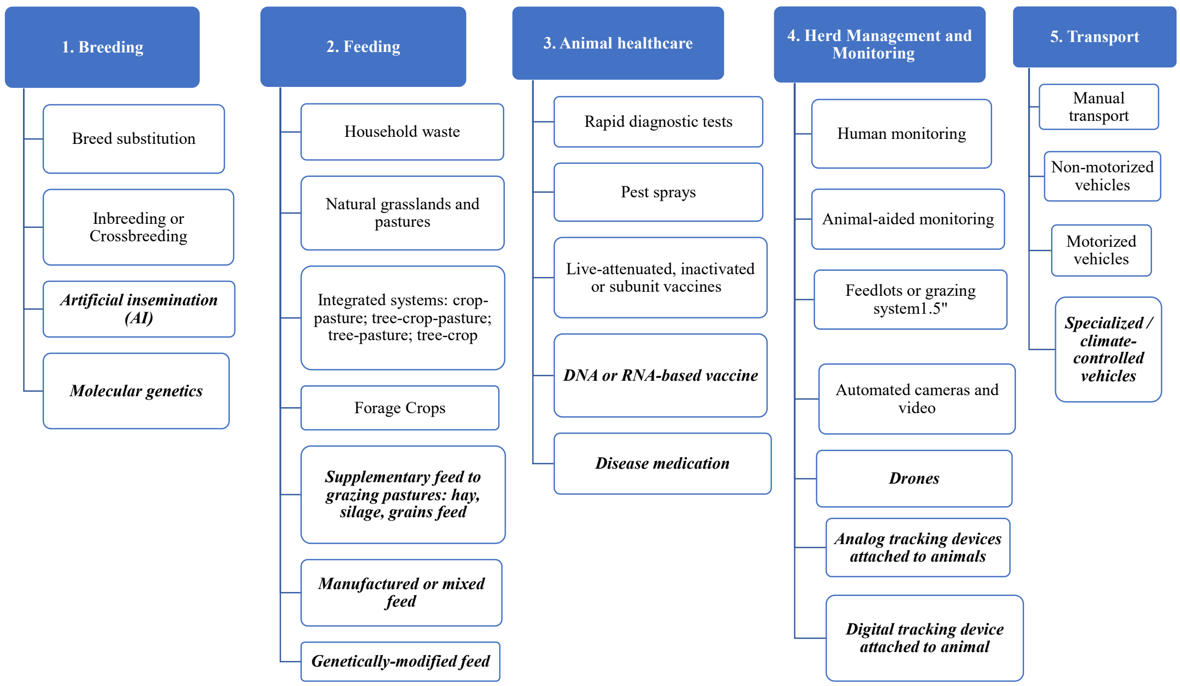


Figure A.1: Agriculture - Livestock: Business Functions and Technologies

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

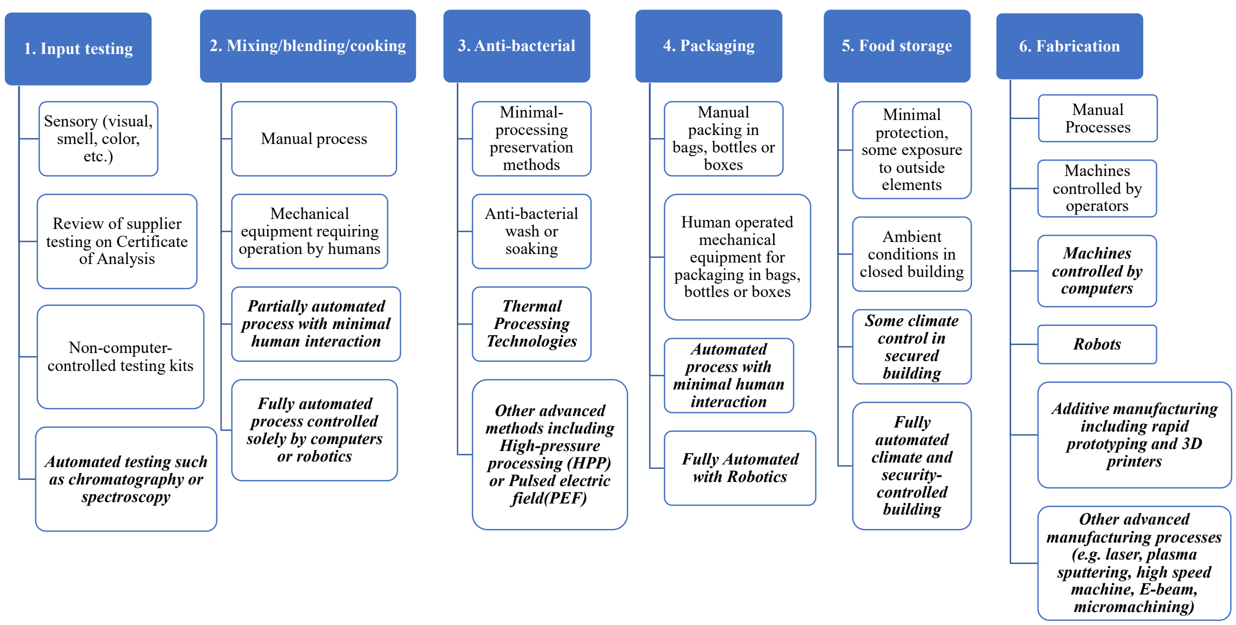


Figure A.2: Food Processing: Business Functions and Technologies

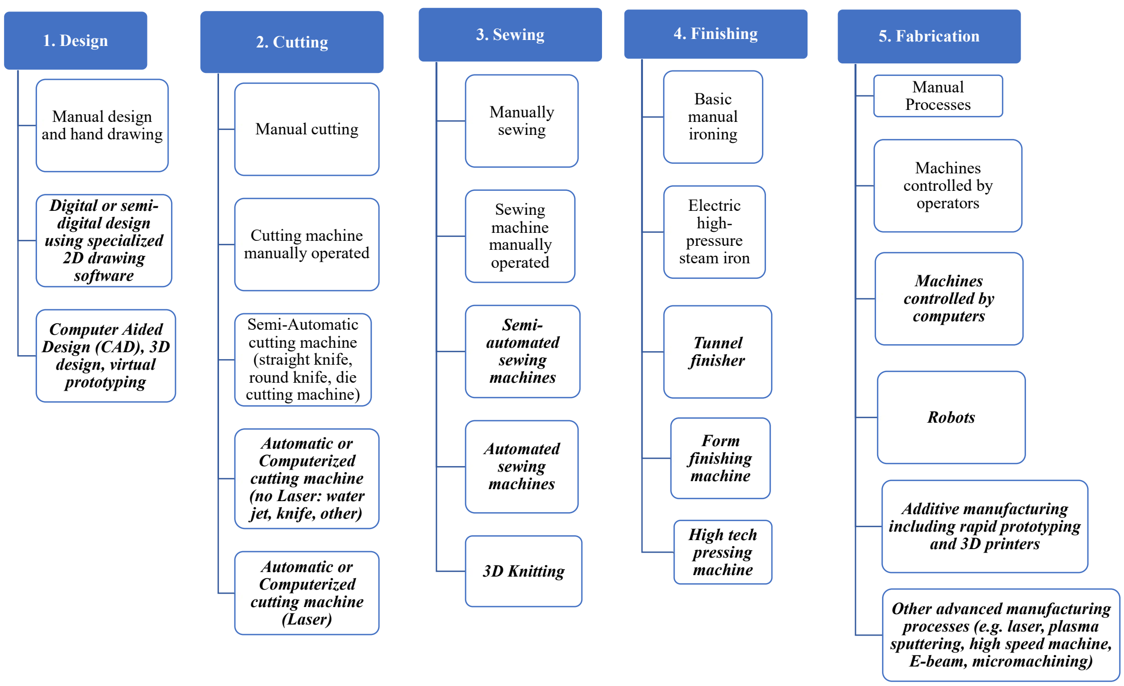


Figure A.3: Wearing Apparel: Business Functions and Technologies

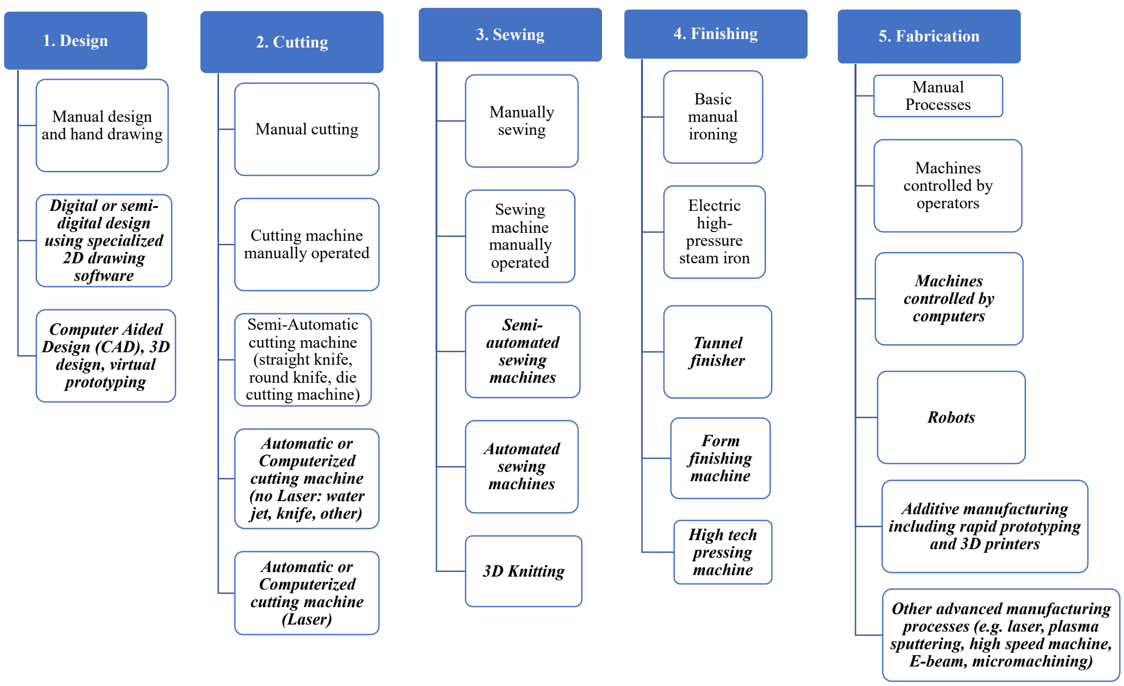


Figure A.4: Leather and Footwear: Business Functions and Technologies

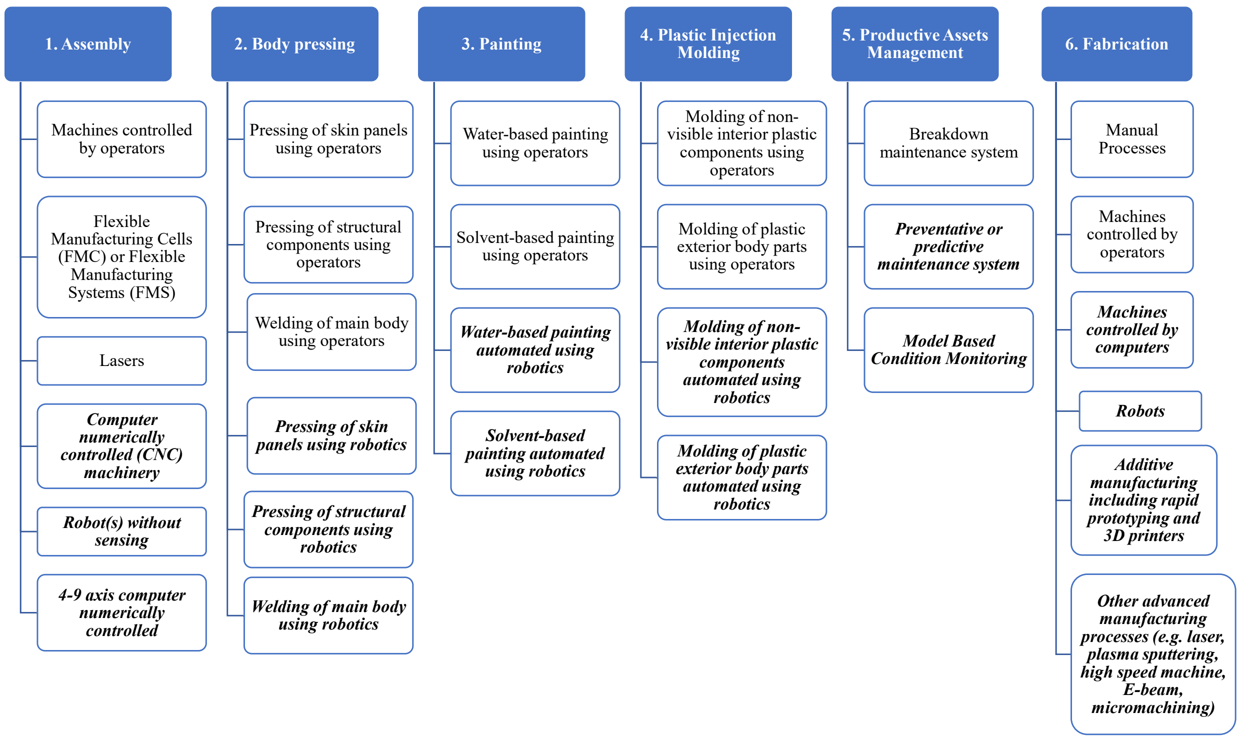


Figure A.5: Automotive: Business Functions and Technologies

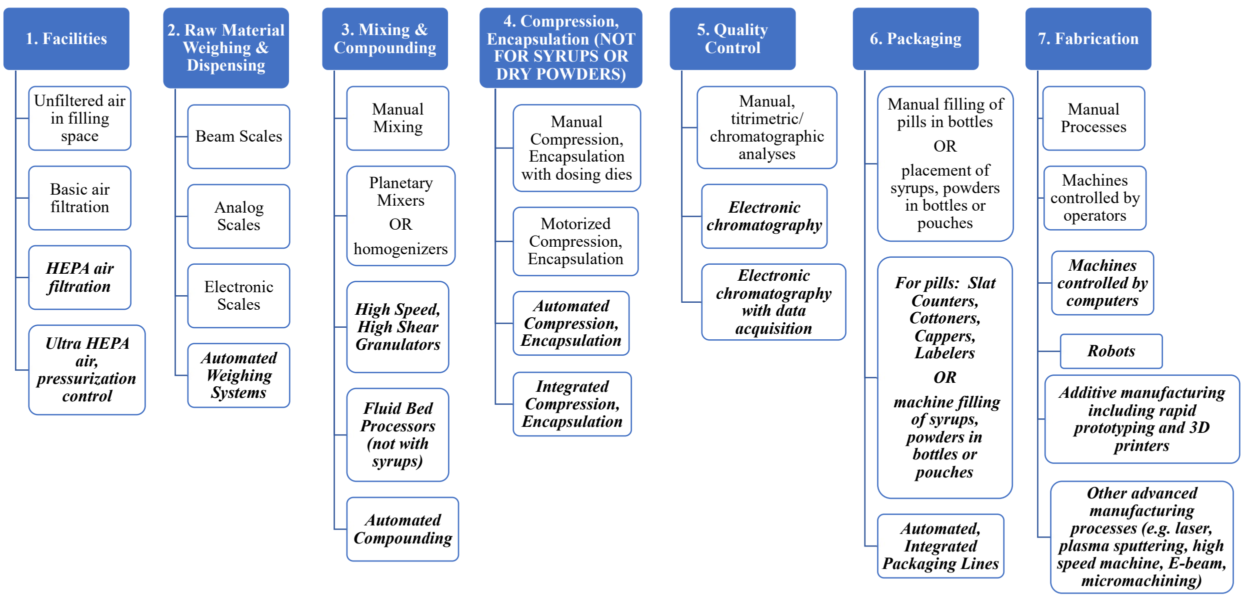


Figure A.6: Pharmaceutical: Business Functions and Technologies

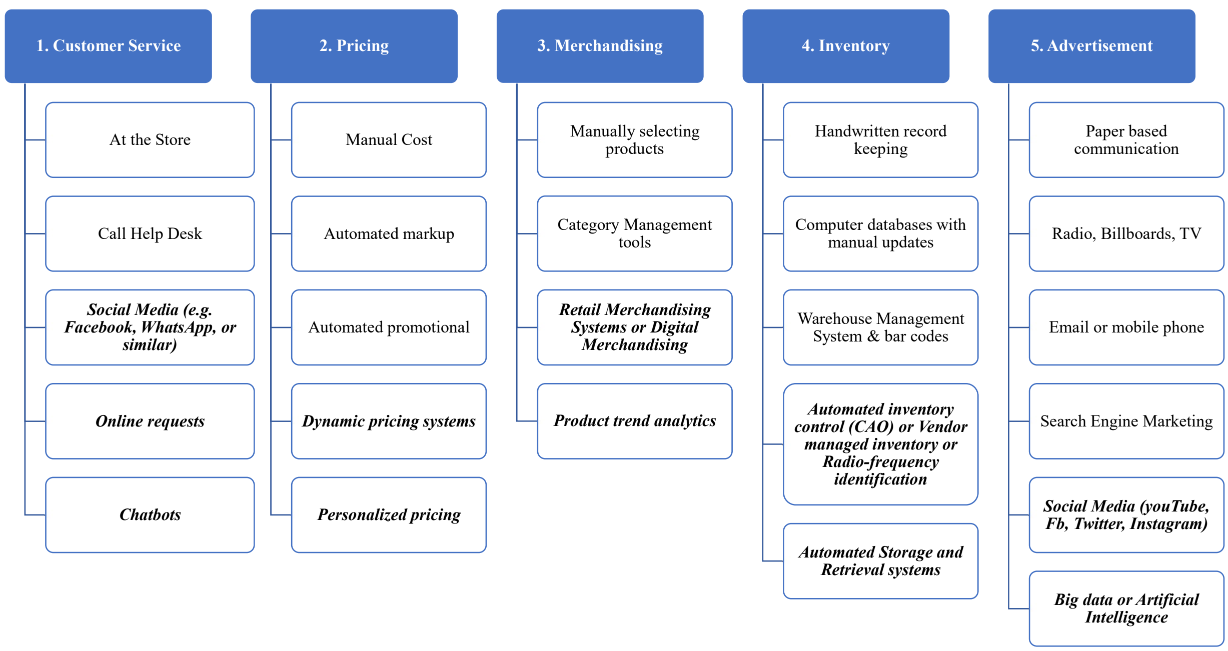


Figure A.7: Wholesale and Retail: Business Functions and Technologies

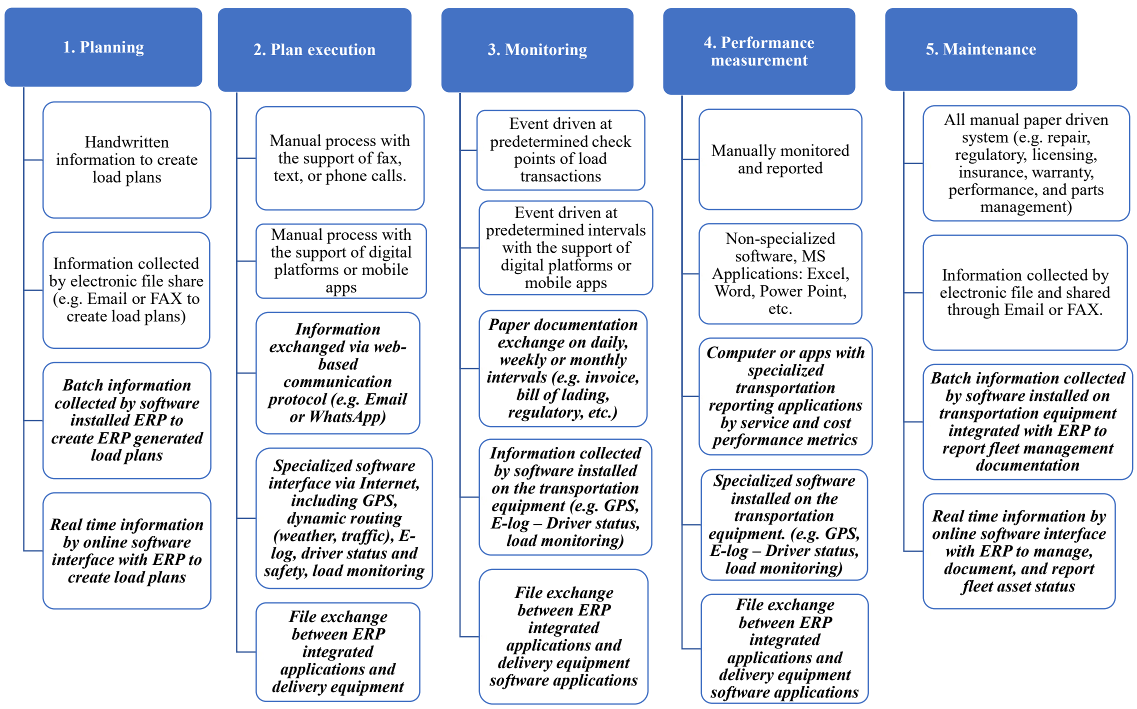


Figure A.8: Land Transportation: Business Functions and Technologies

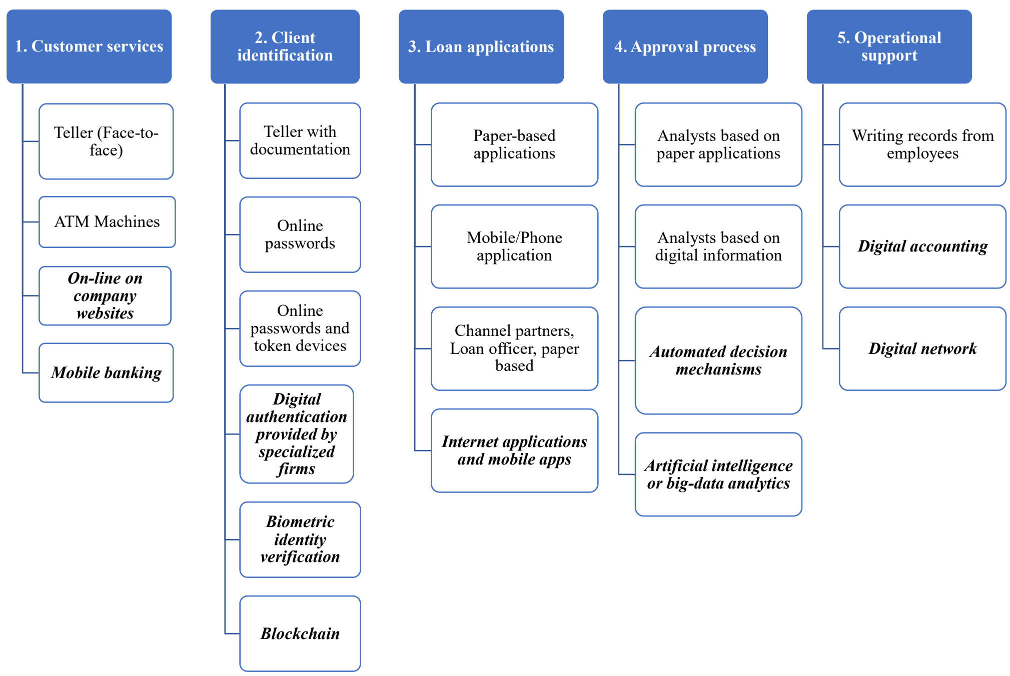


Figure A.9: Financial Services: Business Functions and Technologies

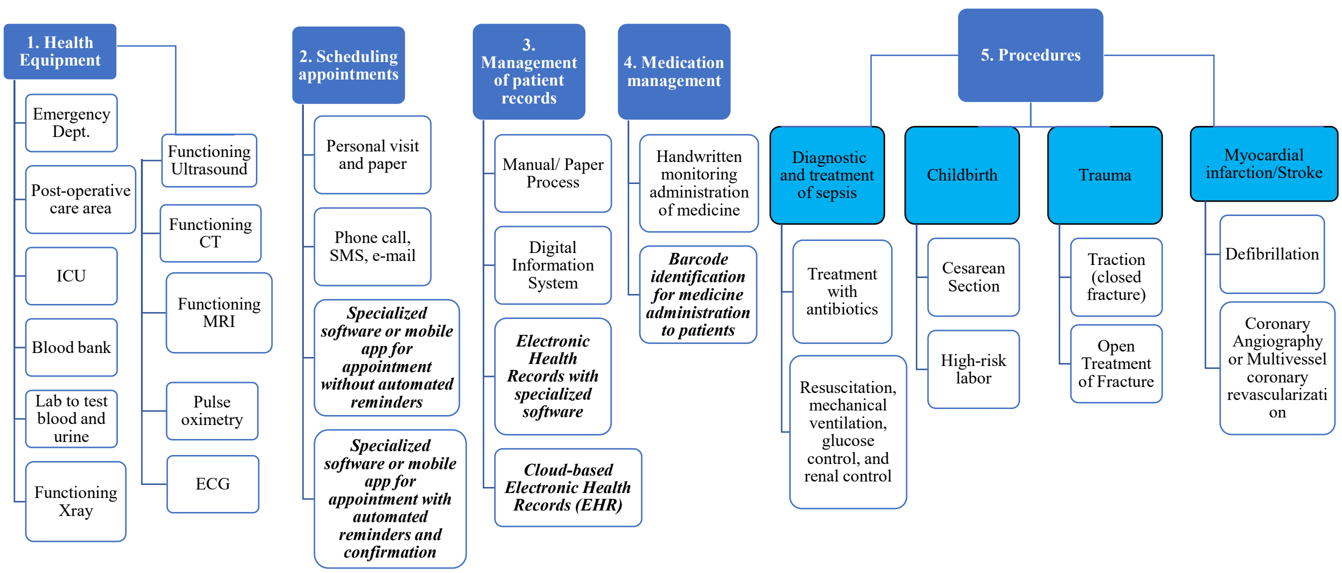


Figure A.10: Health Services: Business Functions and Technologies

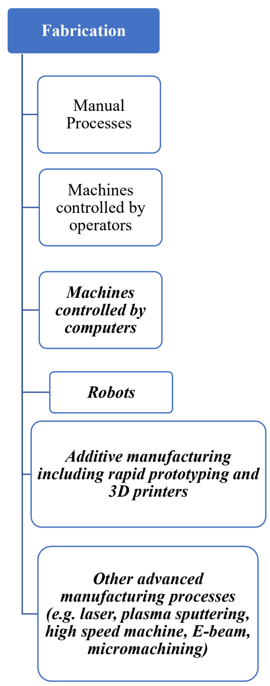


Figure A.11: Other Manufacturing: Business Functions and Technologies

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.[29]

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;

[29] 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 Financial Statements and Workers

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

Financial statements. 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.

Workers. 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 Validation of the Ranking of Technology Sophistication

To evaluate the coherence of the rankings of technology sophisticated from industry experts, we implement a multi-stage validation exercise. First, we select 14 business functions for which we compare the technologies in the grid along three dimensions that define their sophistication and that summarize some of the arguments used by experts to justify their rankings: functionality, integration, and automation. Functionality refers to the capabilities a technology offers to handle more complex tasks, in a faster way, on a larger scale, with greater accuracy and reliability. Integration reflects a technology's ability to connect and interact seamlessly with other systems by exchanging data and coordinating processes. Automation enables the technology to execute processes, make decisions, and generate outcomes independently, without human intervention.

For these functions, we also document the year of the launch and the cost of each of the technologies in the grid. Though these variables do not define the sophistication of a technology per se, more sophisticated technologies tend to have been developed more recently and be more expensive. This is particularly the case for sector specific technologies used in the production of goods (e.g., agriculture and manufacturing).

We relied on multiple source collect information on the functionality, integration, and automation features, as well as the launch and the cost of the technologies in the grid. In addition to broad desk research, we consult official websites with description of specific leading brands supplying these technologies (e.g., Microsoft website provides detailed information on the features of Microsoft Excel, including prices, which is used as a *proxy* for standard spreadsheet software, one of the technologies in the grid for business administration).

In addition, we validate the rankings by using AI-powered large language models. First, we asked ChatGPT to rank the technologies in the grid in terms of technology sophistication. Second, we ask ChatGPT to rank technology sophistication based on specific definitions of functionality, integration, and automation and compare those with the industry experts rankings resulted in the technology grid, based on the methodology we described in section 2. Third, we ask ChatGPT to identify a specific task for each of the 14 business functions and estimate the time required to perform the task with each technology.

Tables A.1 - A.13 summarize these exercises for the following business functions: sourcing, marketing, sales, payment, and quality control, among GBFs; Weeding and harvesting for agriculture; Design and sewing for apparel; Anti-bacterial and packaging for food processing; Pricing and inventory for retail services; in addition to the example for businesses administration, provided in section 2. We also added for each business function a summary of two of the exercises with ChatGPT: i) The overall ranking of technology sophistication; ii) The estimated time to perform a typical task in the business function. Next sub-section provides more details on these and additional exercises conducted with ChatGPT. Subsection A.2.1 describes the specific ChatGPT prompts used for these exercises and displays scatter plots showing the positive association between ChatGPT and industry expert rankings of technology sophistication.

Overall, the validation exercise across 14 businesses functions strongly supports the experts original ranking of sophistication. To start, within each business function, the specific features of technologies are associated with functionality, integration, and automation.

Table A.1: Technology Sophistication in Sourcing

	Manual Search	Computers with Standard Software	Mobile Apps or Digital Platforms	SRM (not integrated)	SRM (integrated with production planning)
Functionality	Basic functions (e.g., manually finding suppliers in the yellow list)	Moderate features for tracking supplier information and managing procurement database.	Good features for discovering suppliers, comparing options, and managing basic procurement tasks.	Extensive features for supplier management, procurement, contract management, and performance monitoring.	Comprehensive features for supplier management, procurement, contract management, and performance monitoring.
Integration	No.	Limited. It requires manual import/export functions and data entry.	Integrates well with other digital tools, but lacks customization.	Good integration with other procurement tools, but not with production planning.	Seamless integration with production planning and other systems.
Automation	No.	Basic automation through formulas and macros; extensive manual setup and maintenance are required.	Moderate level of automation for tasks such as supplier searches, basic order processing, and communication.	Advanced automation for supplier selection, order processing, and performance tracking.	High level of automation for tasks such as supplier selection, order processing, performance tracking, and demand forecasting.
Rank	1	2	3	4	5
Reason for the Rank	Least sophisticated due to complete lack of integration and automation, relying entirely on manual processes.	Low sophistication due to limited integration and basic automation, mainly suitable for small-scale or manual procurement processes.	Moderate sophistication with decent integration and automation, but mainly focused on task management and basic procurement functions.	Sophisticated in supplier management and procurement with advanced automation, but lacks integration with production planning, limiting its overall functionality.	High level of sophistication with specialized integration for production planning and robust automation, though focused on supplier management and production-related tasks.
Technology Examples	Phone Calls, Personal Visits, Supplier Catalogs	Microsoft Excel (1985), Google Sheets (2006), LibreOffice Calc (2011)	LinkedIn (2003), ThomasNet (1898), Alibaba (1999)	SAP Ariba (1996), Oracle Procurement Cloud (2010), Coupa (2006)	SAP Integrated Business Planning (2014), Oracle SCM Cloud (2016), Infor Nexus (2019)
Cost	Cost of phone calls, travel, accommodation, and catalogs varies	Microsoft Excel: \$150/year, Google Sheets: Free or \$6-\$18/user/month, LibreOffice Calc: Free	LinkedIn: \$59/month, ThomasNet: Free or Custom Pricing, Alibaba: Transaction fees (varies)	SAP Ariba: Pricing depends on enterprise size; Oracle Procurement Cloud: Starts at \$625/user/month, Coupa: \$2,500/month	SAP IBP: Starts at \$30,600/year; Oracle SCM Cloud: Starts at \$350/month; Infor Nexus: Starts at \$10,000
Launch Year	N/A	1985, 2006, 2011	2003, 1898, 1999	1996, 2010, 2006	2014, 2016, 2019
ChatGPT ranking	1	2	3.5	4	5
Estimated time in a task	4 hours	2.5 hours	1.5 hour	1 hour	30 minutes

Source: The table draws upon various sources of information including desk research, website of various companies including Microsoft, Google, LinkedIn, ThomasNet, Alibaba, Oracle, Coupa, InforNexus, and SAP. *CHATGPT estimates the time for selecting suppliers and managing orders, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.2: Technology Sophistication in Marketing

	Face-to-face chat	**Online chat**	**Structured surveys**	**Customer Relationship Management software (CRM)**	**Big Data Analytics or Machine Learning algorithms (AI)**
Functionality	Collecting qualitative data, gaining in-depth insights, building rapport with customers.	Real-time communication with customers, collecting feedback, handling customer queries.	Designing surveys, collecting structured feedback, analyzing responses.	Managing customer data, tracking interactions, automating marketing campaigns, analyzing customer behavior.	Advanced analytics, machine learning, predictive modeling, real-time data processing.
Integration	No integration	Basic integration with CRM systems and data analysis tools.	Moderate integration with data analysis tools and CRM systems.	Good integration with other business systems, including email marketing, social media, and sales platforms.	Seamless integration with multiple data sources, CRM systems, and marketing platforms.
Automation	No automation	Basic automation for responses and data collection; advanced features available in some tools.	Moderate level of automation for survey distribution and response collection, but manual analysis is often required.	High level of automation for tracking interactions, segmenting customers, and running marketing campaigns.	High level of automation for data collection, analysis, and generating insights.
Rank	5	4	3	2	1
Reason for Ranking	Least sophisticated due to the absence of integration and automation, although valuable for qualitative insights.	Effective for immediate customer interaction but lacks the sophistication in automation and integration of higher ranks.	Useful for structured feedback, but lower automation and integration limit its sophistication compared to top ranks.	Strong integration and automation, making it highly effective for comprehensive customer management and marketing.	Highest sophistication due to advanced analytics, seamless integration, and high automation for actionable insights.
Technology Examples	In-person Interviews, Focus Groups, Mystery Shopping	WhatsApp Business, Intercom, LiveChat	SurveyMonkey, Qualtrics, Google Forms	Salesforce, HubSpot CRM, Zoho CRM	Google Analytics
Launch Year	N/A	2018 (WhatsApp Business), 2011 (Intercom), 2002 (LiveChat)	1999 (SurveyMonkey), 2002 (Qualtrics), 2008 (Google Forms)	1999 (Salesforce), 2014 (HubSpot CRM), 2005 (Zoho CRM)	2005
Cost	Cost of time and travel (varies), \$2,000-\$5,000 per session (Focus Groups), \$20-\$100 per shop (Mystery Shopping)	Free or \$5/month (WhatsApp Business), \$39-\$139/seat/month (Intercom), \$20-\$59/user/month (LiveChat)	Free or \$25-\$75/user/month (SurveyMonkey), Based on invitation (Qualtrics), Free or up to \$18/user/month (Google Forms)	\$1,000-\$10,000/month (Salesforce), \$1,170-\$4,300/month (HubSpot CRM), Free or \$20-\$50/user/month (Zoho CRM)	Free or \$150,000/year
ChatGPT ranking	1	2	3	4	5
Estimated time in a task	4 hours	2 hours	1.5 hours	1 hour	30 minutes

Source: The table draws upon various sources of information including desk research, and websites of various companies including Google, Salesforce, Hubspot, Zoho, SurveyMonkey, Qualtrics, WhatsApp, Intercom, and LiveChat. *CHATGPT estimates the time for collecting customer feedback, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.3: Technology Sophistication in Sales

	Sales at the establishment's premises	Sales by phone, email orders, or sales representatives	Sales through social platforms	Online sales using external digital platforms	Online sales (own website)	Electronic orders integrated to SCM systems
Functionality	Basic features for point of sale, inventory management, and sales tracking.	Features for managing customer data, tracking sales activities, and automating sales processes.	Basic features for creating online storefronts, managing product listings, and processing sales orders.	Features for listing products, managing inventory, processing orders, and handling shipping and returns.	Extensive features for creating and managing online stores, processing payments, handling shipping, and inventory management.	Comprehensive features for managing e-commerce, order processing, inventory management, and supply chain integration.
Integration	Limited integration with other systems.	No integration with CRM and email systems.	Limited integration with inventory management and processing systems.	Good integration with inventory management and shipping services.	Good integration with payment gateways, shipping carriers, and inventory management systems.	Seamless integration with supply chain, inventory, and logistics systems.
Automation	Basic automation for sales tracking and inventory updates.	Moderate automation for tracking interactions and managing sales activities.	Moderate level of automation for order processing and inventory updates.	Moderate level of automation for order processing and inventory updates.	High level of automation for order processing, payment handling, and inventory updates.	High level of automation for order processing, inventory updates, and supply chain management.
Rank	6	5	4	3	2	1
Reason	Basic functionality and minimal integration or automation, suitable for traditional retail.	Limited sophistication due to lack of integration, with moderate automation focusing on sales tracking.	Moderately sophisticated with strong integration and moderate automation, suitable for third-party selling.	High sophistication with extensive e-commerce functionality, good integration, and significant automation.	High sophistication with e-commerce integration and automation	Most sophisticated due to its comprehensive functionality, seamless integration with supply chains, and high automation.
Technology Example	Offline stores	WhatsApp/ Email/ Salesmen	Facebook Shops	Amazon Seller Central	Shopify	SAP Commerce Cloud
Cost	Cost of setting up stores	Cost of mobile, internet, travel	Facebook Shops	$39.99/month + selling fees	$29-$299/month	Starts at $100,000/year
Launch Year	n/a	n/a	2020	2000	2006	2013
ChatGPT ranking	1	2	3	4	4.5	5
Estimated time in a task	3 hours	2 hours	1.5 hours	1 hour	30 minutes	30 minutes

Source: The table draws upon various sources of information including desk research, and websites of various companies including SAP, Shopify, Amazon, Facebook, and WhatsApp. *CHATGPT estimates the time for processing sales orders, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.4: Technology Sophistication in Payment

Criteria	Exchange of goods (Barter Exchange Platforms)	Cash(Physical Currency)	Cheque, voucher, or bank wire at the branch (Bank Cheque)	Prepaid card, Debit card, or Credit card (Visa)	Online or electronic payment through a bank wire (SWIFT)	Online payment through platform (PayPal)	Virtual or Cryptocurrency (Bitcoin)
Functionality	Facilitating the direct exchange of goods and services without using money	Direct payment using coins and banknotes	Payment through written orders to a bank to pay a specific amount from a person's account	Electronic payment method allowing payment directly from a linked bank account or credit line	Secure international financial messaging for transferring funds between banks	Online payment system for sending and receiving money, purchasing goods and services	Decentralized digital currency for peer-to-peer transactions
Integration	No integration	No integration	Basic integration with banking systems for processing cheques	Wide integration with merchants, banks, and payment processors	Extensive integration with global banking systems	Seamless integration with e-commerce platforms, banking systems, and other financial tools	Good integration with digital wallets, exchanges, and some e-commerce platforms
Automation	No automation	No automation	Limited automation; manual processing required	High level of automation for transaction approval and record-keeping	High level of automation for fund transfers and transaction tracking	High level of automation for transaction processing, payment notifications, and record-keeping	High level of automation for transaction validation and record-keeping via blockchain technology
Rank	7	6	5	4	3	2	1
Reason for Ranking	Basic functionality with no integration or automation, the least sophisticated method in the context of payments	Simple functionality with no integration or automation, making it the least sophisticated	Limited functionality and integration, with minimal automation, making it less sophisticated than newer methods	Widely accepted with robust integration and automation, although slightly less sophisticated than online platforms	Strong integration within global banking, with reliable automation, but more suited for large transactions	High functionality, broad integration across multiple platforms, and high automation in processing payments	High automation, but integration is less widespread and functionality is limited to certain platforms
Average Costs	Transaction fees, varies	No cost	Bank fees, varies	Merchant fees (1.5%-3%), annual fees ($0-$100)	Bank fees, varies	2.9% + $0.30 per transaction	Transaction fees (varies), network fees
Launch Year	N/A	N/A	N/A	1958	1977	1998	2009
ChatGPT ranking	1	2	2.5	3.5	4	4.5	5
Estimate time in a task 4 hours	2.5 hours	2 hours	1 hour	30 minutes	30 minutes	12 minutes	

Source: The table draws upon various sources of information including desk research, and websites of various companies including Bitcoin, PayPal, SWIFT, and VISA. *CHATGPT estimates the time for processing payments, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.5: Technology sophistication in Quality Control

Category	**Manual, Visual or Written Processes**	**Human Inspection with Computers**	**Statistical Process Control**	**Automated Systems for Inspection**
Functionality	Documenting inspection results, visual checks, manual recording of defects.	Digitizing inspection processes, capturing photos, real-time data entry.	Analyzing data, statistical process control, generating control charts and reports.	Automated visual inspection, defect detection, measurement, sorting.
Integration	No integration	Moderate integration with data management systems and reporting tools.	Good integration with data collection systems and reporting tools.	High integration with manufacturing systems, data collection, and reporting tools.
Automation	No automation	Moderate level of automation for data entry and reporting; inspections still manual.	High level of automation for data analysis, control chart generation, and reporting.	High level of automation for real-time inspection, defect detection, and data logging.
Rank	4	3	2	1
Reason for Ranking	Lowest sophistication as it lacks automation and integration, relying entirely on manual processes.	Moderately sophisticated with digitized processes but less automated inspection; integration is less robust.	Highly sophisticated with automated data analysis, good integration with data systems, but less real-time functionality compared to automated systems.	Highest sophistication due to full automation and advanced integration with manufacturing systems for real-time inspection and data logging.
Technology Example	Checklists and Written Reports	Inspectorio	Minitab	Cognex In-Sight
Average Costs	Cost of paper and printing, varies	Varies	$1,500-$4,000/license	$5,000-$20,000/system
Launch Year	N/A	2016	1972	2016
ChatGPT ranking	1	2.5	3	4
Estimated time in a task	5 hours	2.5 hours	1 hour	30 minutes

Source: The table draws upon various sources of information including desk research, and websites of various companies including Cognex In-Sight, and Minitab, Inspectorio. *CHATGPT estimates the time for inspecting product quality, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.6: Technology sophistication in Agriculture - Weeding

	Manual application of herbicide, fertilizer, or pesticide	**Biological methods of fertilizing, weed, or pest control**	**Mechanical application of herbicide, fertilizer, or pesticide**	**Fully-automated VRA tools (Precision Agriculture)**	**Drone Application (Advanced Precision Agriculture)**
Functionality	Simple tools for applying inputs by hand.	Utilizes natural predators and organic materials to enhance soil fertility and control pests.	Mechanized sprayers can cover large areas and apply inputs consistently.	VRA tools adjust the number of inputs based on real-time soil and plant conditions.	Drones can perform a wide range of tasks such as applying inputs, monitoring fields, and collecting data.
Integration	No integration with digital or mechanical systems.	Minimal integration with other systems; relies on natural processes.	Limited integration with digital technologies; primarily mechanical.	Integrated with soil and plant sensors to provide data-driven application rates.	High integration with remote sensing technology and on-site sensors, providing real-time data and analytics.
Automation	Entirely manual; requires significant labor and time.	No automation; requires human oversight and management.	Requires human operation but significantly reduces the physical effort and time.	Automated application based on sensor data, ensuring optimal use of resources.	Highly automated with minimal human intervention, allowing for precise application and monitoring.
Rank	5	4	3	2	1
Reason for Rank	Least sophisticated with basic functionality, no integration, and no automation.	Low sophistication with natural functionality, minimal integration, and no automation.	Moderate sophistication with effective functionality, but limited integration with digital systems and requires human operation.	High sophistication with precise functionality and good integration with digital systems, but slightly less automated and versatile than drones.	Most sophisticated with high functionality, full integration with digital systems, and high automation.
Technology Example	Hand-held Sprayer	Composting and Natural Predators	Tractor-mounted Sprayer	VRA Sprayer with Sensors	Agricultural Drone with Sensors
Average Costs	$20-$100	$50-$500	$1,000-$10,000	$20,000-$50,000	$10,000-$30,000
Launch Year	N/A (Ancient Tool)	N/A (Ancient Practice)	1950s	2000s	2010s
ChatGPT ranking	1	2.5	3	4.5	5
Estimated time in a task	8-10 hours	4-5 hours	6-7 hours	2-3 hours	1-2 hours

Source: The table draws upon various sources of information including desk research, and websites of various online marketplaces for agricultural tools like Home Depot and Rogers Sprayers. *CHATGPT estimates the time for weeding and pest control in a medium-sized field, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.7: Technology sophistication in Agriculture - Harvesting

	Manual Harvesting, Training, Pruning, or Picking	**Animal Aided Instruments**	**Human-Operated Machines or Single-Axle Tractor (One Specific Function)**	**Mechanized Process (Fully Operated by Worker)**	**Automated Process (Supported by Digital Technologies)**
Functionality	Basic manual tasks (collecting, cutting, pruning).	Performs basic functions with animal assistance (e.g., tilling, transporting).	Performs a specific function (e.g., soil preparation, grass cutting).	Combines cutting, threshing, and cleaning into one operation.	Combines multiple functions (cutting, threshing, cleaning) with high precision.
Integration	No integration, entirely manual.	Limited integration, relies on animal power.	Limited to single-function tasks.	Mechanically integrates multiple harvesting functions.	Integrates GPS and computing systems for optimized paths and resource use.
Automation	No automation, fully reliant on human labor.	No automation, dependent on human and animal labor.	Manually operated, no automation.	Operated manually by the worker, but mechanized.	High level of automation, requiring minimal human intervention.
Rank	5	4	3	2	1
Reason for Rank	Least sophisticated with basic functionality, no integration, and entirely manual labor.	Low sophistication, relying on animal labor with no digital integration or automation.	Moderate sophistication with effective functionality, but limited to single functions and requires full manual operation.	High functionality with integration of multiple mechanical functions, but requires manual operation.	Most sophisticated with advanced functionality, full integration with digital systems, and high automation.
Technology Examples	Harvesting Bags, Felco Pruning Shears, Hori Hori Knife	Horse-Drawn Reaper, Ox-Drawn Mower, Animal-Powered Binder	Walk-Behind Sickle Bar Mower, Single-Axle Mower, Hand-Pushed Seeder	Tractor-Mounted Harvester, Combine Harvester, Self-Propelled Forage Harvester	GPS-Enabled Tractor with Harvesting Attachment, Auto-Steer Combine Harvester, Smart Forage Harvester
Average Costs	\$30, \$50, \$35	\$300, \$250, \$400	\$600, \$600, \$450	\$7,000, \$300,000, \$150,000	\$200,000, \$400,000, \$350,000
Launch Years	1960, 1945, 1950	1800s, 1800s, 1850s	1960s, 1970s, 1960s	1920s, 1930s, 1940s	1990s, 2000s, 2010s
ChatGPT ranking	1	2	3.5	4	5
Estimated time in a task	200 hours	50 hours	30 hours	2 hours	1.5 hours

Source: The table draws upon various sources of information including desk research, and websites of various online sellers of agricultural equipment including John Deere, CLAAS, and Amazon Marketplace. *CHATGPT estimates the time for harvesting a 1-acre field of wheat, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.8: Technology sophistication in Apparel - Design

	Manual design and hand drawing	**Digital or semi-digital design using specialized 2D drawing software**	**Computer Aided Design (CAD), 3D design, virtual prototyping**
Functionality	Fundamental capabilities for drawing and drafting, relying on user skill and precision	Robust capabilities for vector graphics, layout design, and technical drawing	Comprehensive 2D and 3D design features, including advanced modeling, simulation, and rendering
Integration	No digital integration; standalone and physical	Limited integration; can import/export various file types	High integration with CAM, CAE, and other tools for seamless workflow
Automation	No automation; all aspects are manual	Moderate automation with snap-to-grid, alignment tools, and reusable components	High level of automation with automatic dimensioning, constraint management, and generative design
Rank	3	2	1
Reason for Ranking	Lowest sophistication due to reliance on manual input, lack of integration, and absence of automation	Moderate sophistication with strong 2D design capabilities, but lower integration and automation compared to CAD tools	Highest sophistication due to advanced functionality, seamless integration, and high automation, ideal for complex apparel design
Technology Examples	Drafting Table, Mechanical Pencil, Drawing Paper	Adobe Illustrator, CorelDRAW, AutoCAD LT	AutoCAD, SolidWorks, Autodesk Fusion 360
Cost	\$2 - \$300	\$20.99/month - \$499 (one-time) or \$198/year	\$60/month - \$3,995 (one-time) + \$1,295/year maintenance
Launch Year	N/A	1987 - 1993	1982 - 2013
ChatGPT ranking	1	3	4.5
Estimated time in a task	8-10 hours	4-6 hours	2-3 hours

Source: The table draws upon various sources of information including desk research, and websites of companies like Autodesk, SolidWorks, CorelDraw, and Adobe.
*CHATGPT estimates the time for designing a simple dress, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.9: Technology sophistication in Apparel - Sewing

	Manually Sewing	**Machine Sewing Manually Operated**	**Semi-automated Sewing Machines**	**Automated Sewing Machines**	**3D Knitting**
Functionality	Basic hand stitching, mending, and precise cutting of fabric.	Basic stitching with manual adjustments for tension and stitch length.	Automatic tension adjustment, multiple stitch patterns, and built-in embroidery designs.	High-speed sewing, automated thread trimming, and programmable stitch patterns.	Knits entire garments in 3D with no seams, enabling complex designs and high customization.
Integration	No integration	Limited to no integration with digital or automated systems.	Integrated with pre-programmed patterns and digital interfaces for ease of use.	Often integrated with manufacturing systems for efficient production workflows.	Integrated with digital design software, allowing for seamless transition from design to production.
Automation	No automation	Minimal automation; relies on user skill and manual adjustments for operation.	Moderate level of automation with features like automatic needle threading and thread cutting.	High level of automation, reducing the need for manual adjustments and interventions.	Fully automated knitting process, from pattern creation to finished garment.
Rank	5	4	3	2	1
Reason for Ranking	Lowest sophistication, entirely manual with no automation or digital integration.	Lower sophistication due to minimal automation and lack of integration with digital systems.	Moderate sophistication with useful automation features, but limited compared to fully automated systems.	High sophistication with advanced functionality and integration, but less specialized than 3D knitting.	Highest sophistication due to full automation, advanced functionality, and seamless integration with digital design tools.
Technology Examples	Needle and Thread, Thimble, Fabric Scissors	Mechanical Sewing Machine, Hand-Crank Sewing Machine, Foot-Pedal Sewing Machine	Electronic Sewing Machine, Computerized Sewing Machine, Embroidery Machine	Industrial Sewing Machine, Automated Quilting Machine, Automated Serger	Shima Seiki WholeGarment, Stoll ADF 3, Kniterate
Cost	$1 - $30	$100 - $500	$300 - $3,000	$1,000 - $20,000	$7,500 - $250,000
Launch Year	Ancient - N/A	1800s - 1900s	1970s - 1990s	1960s - 2000s	1970s - 2018
ChatGPT ranking	1	2.5	3.5	4.5	5
Estimated time in a task	8-10 hours	3-4 hours	2-3 hours	1-2 hours	30 minutes-1 hour

Source: The table draws upon various sources of information including desk research, and websites of companies like Kniterate as well as online sellers of sewing equipment like Amazon Marketplace. *CHATGPT estimates the time for sewing a Simple T-shirt, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.10: Technology sophistication in Food Processing - Anti-Bacterial

	Minimal-processing Preservation	**Anti-bacterial Wash or Soaking**	**Thermal Processing Technologies**	**Other Advanced Methods (HPP, PEF)**
Functionality	Slows bacterial growth without killing pathogens, preserving food quality.	Reduces bacterial load on food surfaces, effective for surface decontamination.	Effective in killing bacteria and pathogens through heat, ensuring food safety.	HPP and PEF offer advanced bacterial inactivation with minimal impact on food quality. UV treatment provides surface-level decontamination.
Integration	Easy to integrate, requires basic modifications to packaging/storage processes.	Simple to integrate into existing systems with easy-to-implement tanks or spray systems.	Well-integrated in the food industry, adaptable to different production scales.	Moderate to high integration into existing production lines with minimal adjustments required.
Automation	Limited automation potential, requiring manual checks and adjustments.	Can be automated but requires some manual intervention for effective application.	Highly automated, offering precise control over processing conditions.	Fully automated processes requiring minimal human intervention.
Rank	4	3	2	1
Explanation for Ranking	Lower sophistication due to minimal bacterial inactivation and reliance on manual processes for maintaining effectiveness.	Moderate sophistication, as it is effective and easy to integrate, but with lower automation levels and more manual intervention needed.	High sophistication due to reliable pathogen inactivation, well-established integration in the industry, and high levels of automation.	Highest sophistication due to advanced, non-thermal bacterial inactivation methods, high integration potential, and full automation.
Technology Examples	Modified Atmosphere Packaging (MAP), Vacuum Packaging, Edible Coatings	Chlorine Dioxide Wash, Organic Acid Wash, Ozone Treatment	Steam Pasteurization, Infrared Heating, Microwave Pasteurization	High-Pressure Processing (HPP), Pulsed Electric Field (PEF), Ultraviolet (UV) Treatment
Cost	$10,000 - $100,000	$500 - $20,000/year	$50,000 - $500,000	$10,000 - $2,500,000
Launch Year	1980s - 2000s	1990s - 2000s	1980s - 2000s	1990s - 2000s
ChatGPT ranking	2	2.5	4	4.5
Estimated time in a task	1-2 hours	15-30 minutes	30 minutes - 1 hour	1-2 hours

Source: The table draws upon various sources of information including desk research, and some lab reports to provide cost estimates of various anti-bacterial processes. *CHATGPT estimates the time for reducing the microbial load on fresh produce, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.11: Technology sophistication in Food Processing - Packaging

	Manual Packing in Bags, Bottles, or Boxes	**Human Operated Mechanical Equipment**	**Power Equipment Requiring Routine Interaction**	**Power Equipment Controlled by Computers/Robotics**
Functionality	Basic functionality involving manual sealing of bags, wrapping of products, and application of labels to packages.	Functionality includes weighing and filling of packages, sealing of bags, and capping of bottles with human-operated mechanical equipment.	High functionality with automated processes for bagging, filling, and wrapping, reducing the need for manual effort.	Highest functionality with advanced capabilities such as precise handling, stacking, and packaging with minimal errors.
Integration	No integration, manual processes not typically integrated with automated systems.	Moderate integration, relatively easy to incorporate into existing workflows but still reliant on human operation.	Moderate to high integration, capable of being incorporated into existing systems but requiring human oversight and routine interaction.	High integration complexity but can seamlessly integrate into existing production lines to enhance efficiency and reduce manual labor.
Automation	No automation, entirely dependent on human labor.	Low to moderate automation, with significant human involvement required for operation and control.	High level of automation, but still requires routine human interaction for loading, monitoring, and occasional adjustments.	Highest level of automation, requiring minimal human interaction once programmed and set up.
Rank	4	3	2	1
Reason for Ranking	Lowest sophistication due to full reliance on manual labor with no automation or integration capabilities.	Moderate sophistication due to reliance on human operation with limited automation, improving speed and accuracy compared to manual methods.	High sophistication due to automated processes with significant human oversight and interaction needed.	Highest sophistication due to advanced functionality, seamless integration, and full automation with minimal human interaction.
Technology Examples	Handheld Heat Sealer, Manual Wrapping Machine, Handheld Label Applicator	Mechanical Weighing and Filling Machine, Foot Pedal Operated Heat Sealer, Semi-Automatic Bottle Capping Machine	Automatic Bagging Machine, Automated Filling Machine, Automated Wrapping Machine	Robotic Palletizer, Automated Guided Vehicles (AGVs), Computer-Controlled Packaging Line
Cost	\$30 - \$300	\$1,000 - \$10,000	\$10,000 - \$100,000	\$50,000 - \$1,000,000
Launch Year	1970s - 1990s	1980s - 2000s	1990s - 2010s	2000s - 2015
ChatGPT ranking	1	2.5	3.5	4.5
Estimated time in a task	8-12 hours	4-6 hours	1-2 hours	0.5-1 hour

Source: The table draws upon various sources of information including desk research, and websites of companies that provide packaging solutions like ZoneSun Auto Pack. *CHATGPT estimates the time for sealing 1,000 Bags of food products, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.12: Technology sophistication in Retail - Pricing

	Manual Cost	**Automated Markup (Excel or similar)**	**Automated Promotional**	**Dynamic Pricing Systems**	**Personalized Pricing driven by Predictive Analytics**
Functionality	Minimal features, relying entirely on manual calculations and record-keeping.	Provides basic automation for pricing calculations, applying markups, and tracking data using formulas and templates.	Focuses on automating promotional pricing based on predefined rules, events, and seasonal factors.	Utilizes AI to optimize prices based on demand prediction, competitive analysis, and real-time data.	Offers the most advanced features, including personalized pricing strategies based on customer behavior, preferences, and predictive analytics.
Integration	No integration	Limited integration capabilities, mainly requiring manual data import/export.	Good integration with sales, marketing, and e-commerce platforms to manage promotions efficiently.	Integrates well with market data sources and internal business systems to provide comprehensive pricing strategies.	Integrates seamlessly with various data sources, CRM systems, and e-commerce platforms to gather comprehensive customer data.
Automation	No automation, entirely dependent on human labor.	Some automation through formulas, but significant manual effort is required for data entry and maintenance.	Automates promotional pricing adjustments, but typically based on predefined schedules and rules rather than real-time data.	Automated price adjustments based on real-time data and algorithms, requiring minimal manual intervention.	Highly automated, using machine learning algorithms to adjust pricing dynamically in real-time based on predictive models and customer data.
Rank	5	4	3	2	1
Reason for Ranking	Lowest sophistication due to the absence of automation and integration, relying entirely on manual calculations.	Lower sophistication with basic automation and limited integration, relying heavily on manual processes for data management.	Moderate sophistication with strong promotional features but limited real-time automation and integration compared to dynamic pricing systems.	High sophistication with strong automation and integration, optimizing prices based on real-time data and AI algorithms.	Highest sophistication due to advanced personalization, seamless integration with multiple platforms, and full automation using AI.
Technology Examples	Pen and Paper	Microsoft Excel, Google Sheets, QuickBooks	Salesforce CPQ, SAP Hybris, Shopify Plus	PROS Pricing, Dynamic Pricing by Prisync, RepricerExpress	Dynamic Yield, Zilliant, Adobe Target
Cost	Minimal to None	$160/year (Office), Free to $12/month, $25/month	$75/user/month, Custom pricing, $2000/month	Custom pricing, $59/month, $55/month	Custom pricing
Launch Year	Pre-digital	1983 - 2006	1997 - 2006	1985 - 2014	1999 - 2011
ChatGPT ranking	1	2.5	3	4	5
Estimated time in a task	8 hours	4 hours	2 hours	1 hour	2 hours

Source: The table draws upon various sources of information including desk research, and websites of companies like Microsoft, Google, Salesforce, SAP, Adobe, PROS, and RepricerExpress. *CHATGPT estimates the time for setting seasonal discount pricing for a product line, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

Table A.13: Technology sophistication in Retail - Inventory

	Handwritten Record Keeping	**Computer Databases with Manual Updates**	**Warehouse Management System**	**Automated Inventory Control**	**Automated Storage and Retrieval Systems (AS/RS)**
Functionality	Basic manual entry and tracking of inventory data.	Basic inventory tracking and data management with manual data entry and updates.	Comprehensive inventory management, order fulfillment, warehouse operations, and supply chain optimization.	Advanced inventory control using RFID for real-time tracking, CAO for automated order replenishment, and VMI for vendor-managed inventory solutions.	Fully automated systems for high-density storage, retrieval, and inventory management.
Integration	No integration	Limited integration capabilities, primarily requiring manual data import/export.	Integrates with ERP, TMS, and other business systems for end-to-end visibility and control.	Integrates with supply chain systems, WMS, and ERP systems for seamless data flow and inventory optimization.	Seamlessly integrates with WMS, ERP, and other business systems to provide real-time data and analytics.
Automation	No automation	Minimal automation, mainly through formulas and templates, with significant manual effort needed.	Offers significant automation for inventory tracking, order processing, and warehouse operations but requires some manual oversight.	High level of automation with real-time tracking, automated updates, and reduced manual intervention.	Provides the highest level of automation, handling storage and retrieval tasks autonomously with minimal human intervention.
Rank	5	4	3	2	1
Reason for Ranking	Lowest sophistication due to the absence of automation and integration, relying entirely on manual processes.	Lower sophistication due to limited automation and integration, with significant manual data entry required.	Moderate sophistication with strong functionality and integration, but with some manual oversight required for operations.	High sophistication due to advanced real-time tracking, automated inventory control, and strong integration with supply chain systems.	Highest sophistication due to full automation, seamless integration, and advanced functionality in high-density storage and retrieval.
Technology Examples	Ledger Books, Paper Forms, Manual Logs	Microsoft Access, Google Sheets, QuickBooks	SAP EWM, Oracle WMS, Manhattan WMS	Zebra Technologies, Blue Yonder, IBM Sterling	Swisslog, Dematic, Honeywell Intelligrated
Cost	Minimal	$160/year (Office), Free to $12/month, $25/month	Custom pricing	Custom pricing	Custom pricing
Launch Year	n/a	1983 - 2006	1990 - 2005	1969 - 2000	1900 - 2001
ChatGPT ranking	1	2	3.5	4	5
Estimated time in a task	40 hours	20 hours	10 hours	4 hours	2 hours

Source: The table draws upon various sources of information including desk research, and websites of companies like Microsoft, Google, Salesforce, SAP, Oracle, Zebra Tech, IBM, SwissLog, Dematic, and Honeywell Integrated. *CHATGPT estimates the time for conducting a full inventory count, following similar characteristics of a given firm. The prompt for these estimates and CHATGPT ranking is available in the appendix.

A.2.1 Comparison between experts' and ChatGPT's sophistication rankings

We validate the industry experts' technology sophistication rankings using AI-powered large language models. To start, we prepared files with the list of business functions and associated technologies in the grid, keeping the way they are described in the FAT survey. We generate separate files for GBFs and each specific sector functions. We then uploaded these files and asked ChatGPT to rank the sophistication of these technologies, providing the following prompts:

1. Based on the survey questions from the document, please create a report ranking the level of technological sophistication for each technology within business function on a scale of 1 to 5, where 1 represents the most basic technology and 5 represents the most sophisticated technology. You can use decimal points. Generate a report file with this information. Provide the data in an Excel sheet.

2. Please expand the previous exercise by providing the rank of technology sophistication for each of these definitions: 1) Functionality: The breadth and depth of features and capabilities that technology offers to perform various tasks and processes within a business function. Higher functionality indicates the ability to handle more complex tasks associated with the business function on a larger scale, faster, with more accuracy and reliability. 2) Integration: The ability of a technology to connect and interact seamlessly with other systems, platforms, or technologies within the business ecosystem. Higher integration means the technology can easily exchange data and coordinate processes with other systems. 3) Automation: The extent to which a technology can perform tasks automatically without human intervention. Higher levels of automation imply that the technology can execute processes, make decisions, and generate outcomes on its own.

3. For each business function, please expand the previous exercise by estimating the time each technology would take to perform a typical task in a company with 10 workers. Please describe this typical task in a separate column and explain the reason for these estimates in another column. Generate a report with the estimates and explanations in Excel.

Tables A.1 - A.13 summarize the main results from prompts 1 and 3. Figures A.12-A.16 show the strong positive association between ChatGPT-specific rankings of technology sophistication for functionality, integration, and automation with industry experts' rankings. The rankings provided by ChatGPT are strongly correlated with our technology sophistication ranking, which is defined by human experts. The overall correlation across all functions and various definitions is above 90%. The results are similar across all business functions.

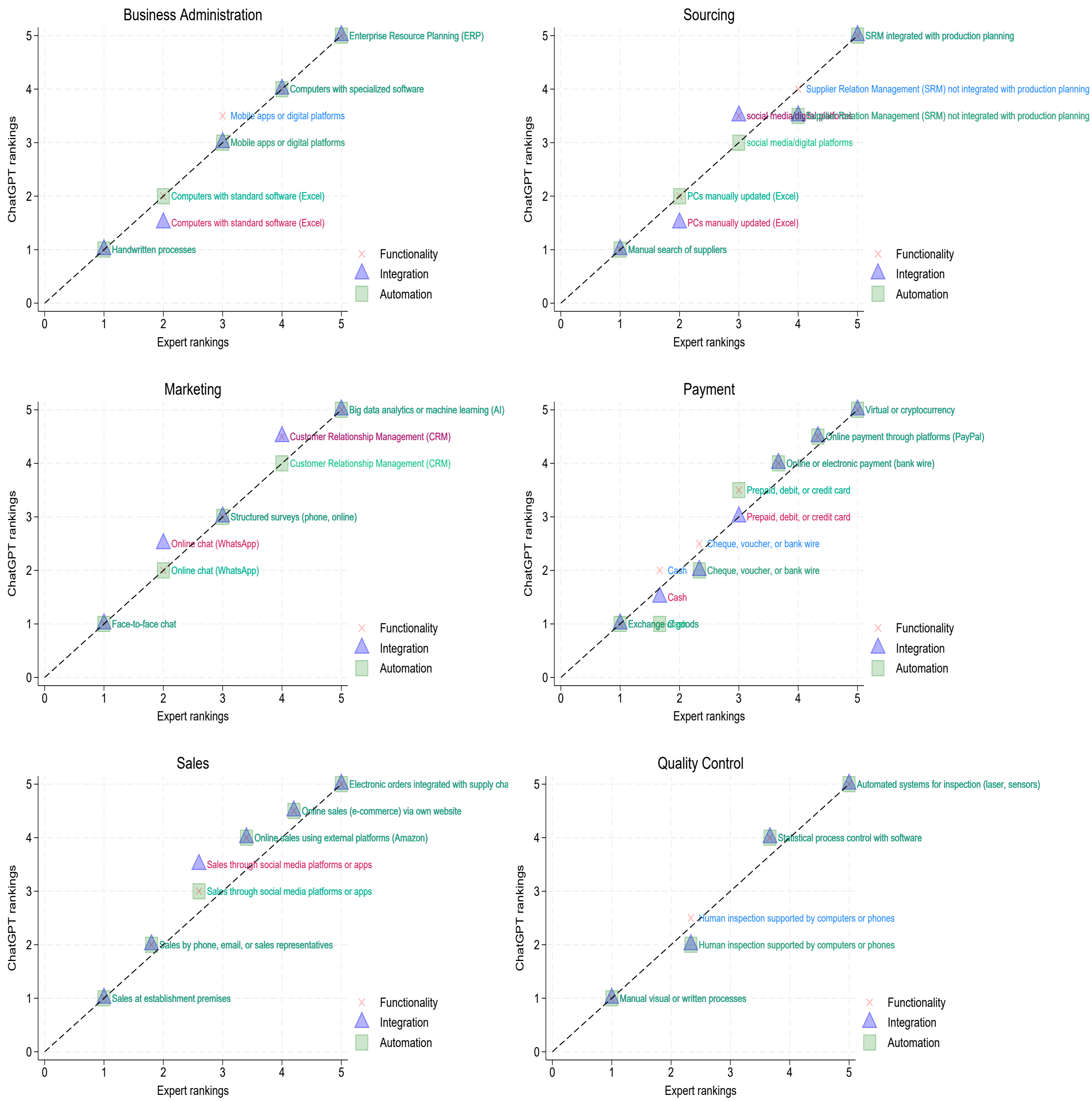


Figure A.12: Comparison between expert's versus ChatGPT's in General Business Functions
Note: This figure compares the experts' rankings of technology sophistication (horizontal axis) with ChatGPT's rankings across the dimensions of functionality, integration, and automation (vertical axis).

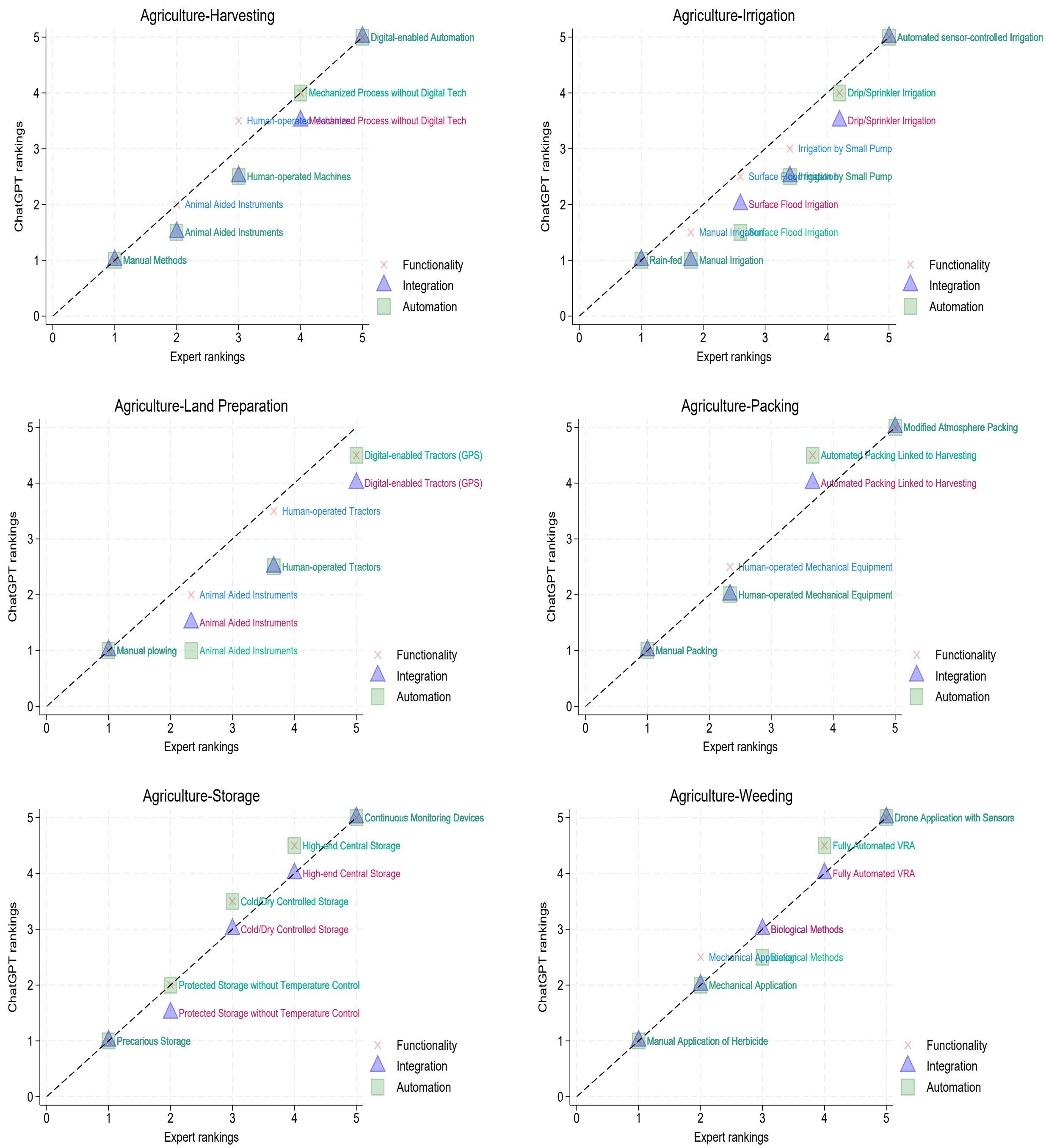


Figure A.13: Comparison between expert's versus ChatGPT's in Agriculture

Note: This figure compares the experts' ranking of technology sophistication (horizontal axis) with ChatGPT's rankings across the dimensions of functionality, integration" "and automation (vertical axis).

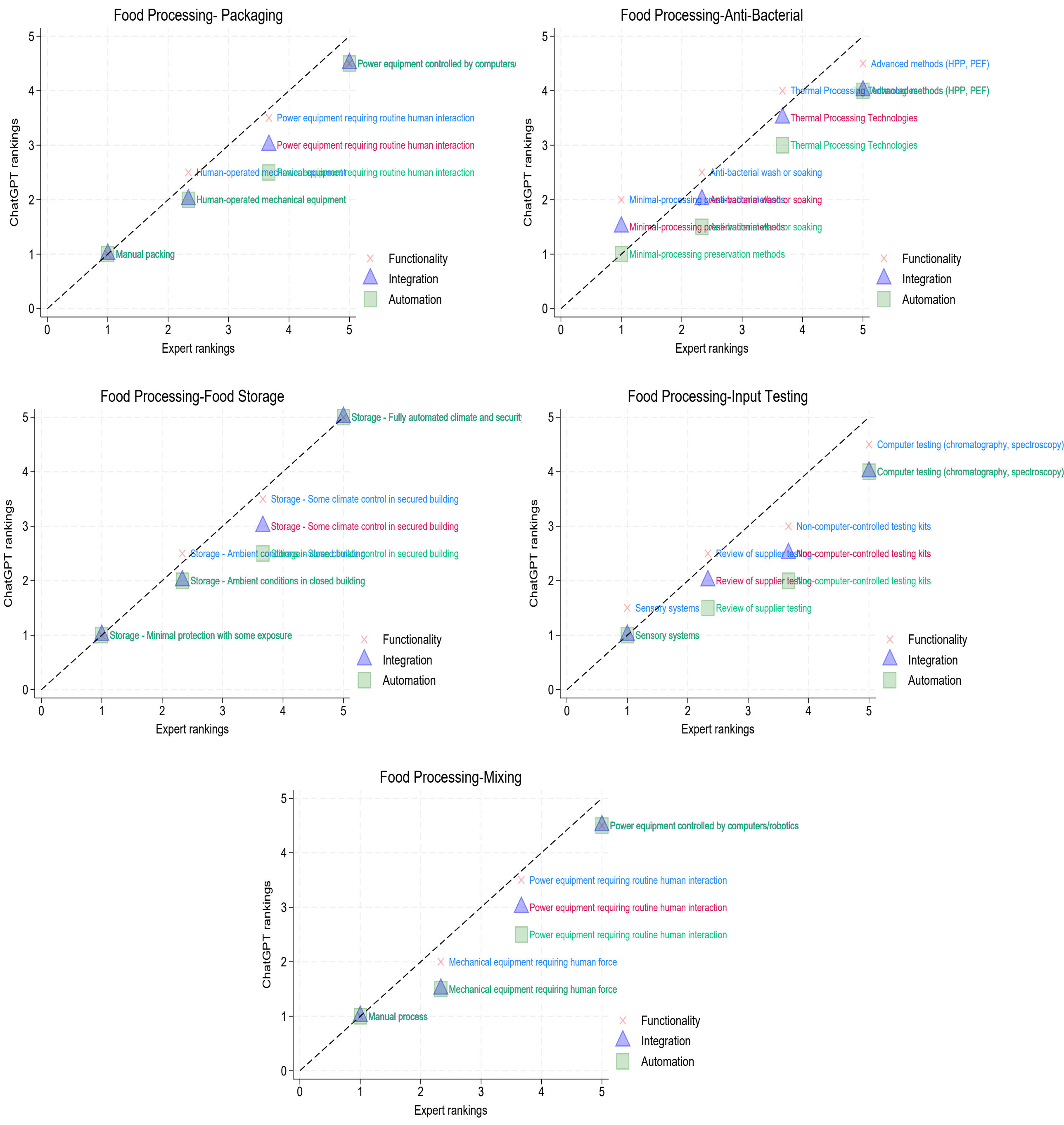


Figure A.14: Comparison between expert's versus ChatGPT's in Food Processing

Note: The figure compares the experts' ranking of technology sophistication (horizontal axis) with ChatGPT's rankings across the dimensions of functionality, integration" "and automation (vertical axis).

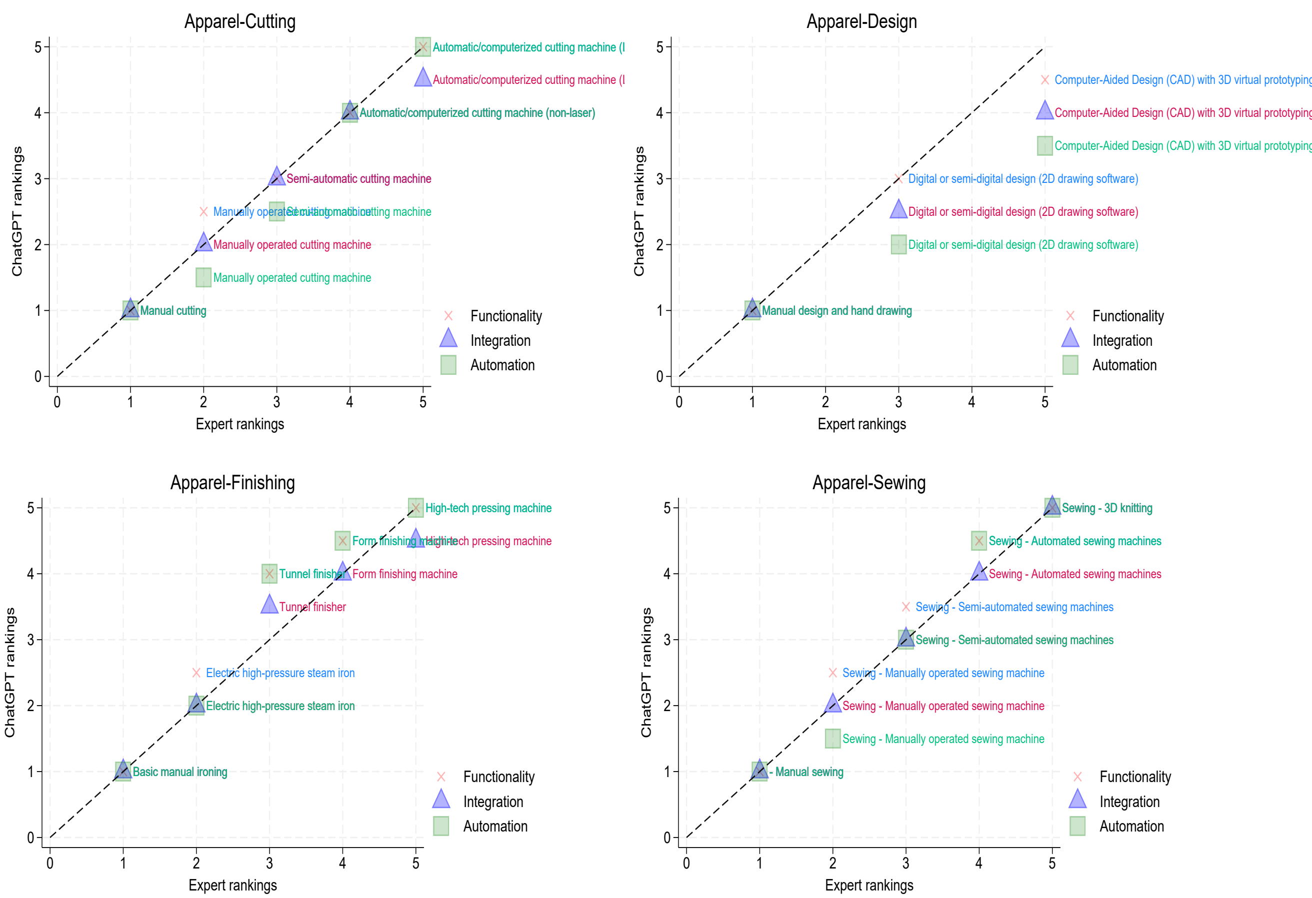


Figure A.15: Comparison between expert's versus ChatGPT's in Apparel

Note: The figure compares the experts' ranking of technology sophistication (horizontal axis) with ChatGPT's rankings across the dimensions of functionality, integration" "and automation (vertical axis).

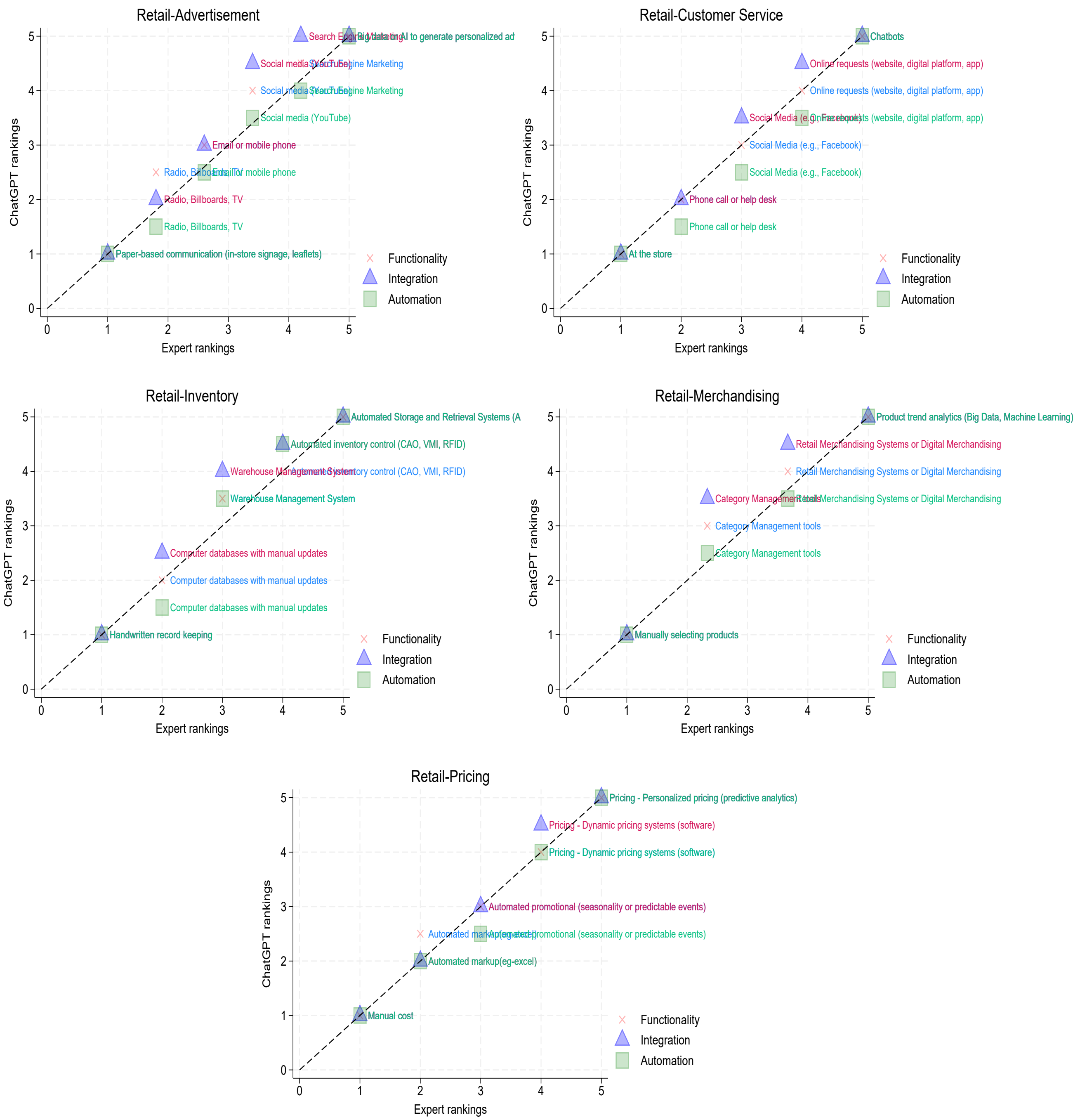


Figure A.16: Comparison between expert's versus ChatGPT's in Retail

Note: The figure compares the experts' ranking of technology sophistication (horizontal axis) with ChatGPT's rankings across the dimensions of functionality, integration" "and automation (vertical axis).

A.3 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.14 provides the main data sources used in the sample frame for each country.

Table A.14: 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.15: 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 2 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.[30] 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), Retail and Wholesale (ISIC 45, 46 and 47), other manufacturing (Group C, excluding food

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

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).[31] 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.16: 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

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

A.4 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.5 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.[32] 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. financial statements' 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

[32] 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,[33] 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.[34] 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

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

[34] 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.[35] 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,

[35]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.6 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.17 shows response rates by country, firm size group and sector. Response rates vary between 15% in Croatia and 86% in Georgia.

Table A.17: 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).[36]

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.18 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.18: 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

[36]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.[37] 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.19 shows that there are no statistically significant differences in technology sophistication between the two groups.

Table A.19: 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.20 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.[38]

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

[37] 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.

[38] See Table A.18 to A.24 in Appendix A.

Table A.20: 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.21 shows that enumerator dummies are not significant for Brazil, Ghana, India, and Korea. For Bangladesh, Senegal, Vietnam, and Kenya, no more than 20% of enumerator dummies are statistically significant with respect to their distribution towards the type of firms they interview across various strata. Table A.22 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

Table A.21: Analysis of enumerator bias distribution

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	0.2
Number of Significantly Different Interviewers	0	0	0	2
Number of Interviewers	44	18	9	10

Note: Data from the Firm-level Adoption of Technology (FAT) surveys in Brazil, Vietnam, Senegal, Bangladesh, Ghana, India, and Korea. 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.

we re-interview 100 randomly selected firms with a short version of the questionnaire. For those firms, we randomly select three business functions and ask about the presence of the relevant technologies.[39] 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.[40] 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

[39]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).

[40]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.22: 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.

compare variables in RAIS with variables we collected in FAT for the same firms.

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.23 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

Table A.23: 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.

firms in RAIS that are part of our universe for the State of Ceará, in Brazil [41]. We then perform a t-test to compare the differences. Table A.24 shows that the differences are not statistically significant.

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

[41] The variables are number of workers, average wages, share of workers with college degree, share of low skilled, and share of high-skilled workers, where high- and low-skilled workers are defined as in Autor and Dorn (2013).

Table A.24: 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 provides details about the construction of the various technology.

B.1 Technology Measures

The technology measures at the business function (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 **??**. To better understand the construction of these variables, it is helpful to understand the structure of the corresponding questions asked. The figure below provides 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.[42]

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.1: Example question for the presence of technologies and most-used technology

If a respondent answers that they use more than one technology in the BF, then they are asked about the most-used technology in that BF.[43]

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

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

After constructing these technology-level binary variables and the variable for most widely-used technology (which are 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 are 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[44].

$ANUM_{f,j}$: $ANUM_{f,j}$ denotes the absolute number of technologies used by an establishment in a specific business function. This variable is calculated by counting the number of technologies in a BF that the establishment confirmed using. Note that $ANUM_{f,j}$ will *always* be less or equal than N_f.

$NUM_{f,j}$: This denotes the relative number of technologies and is calculated using the following formula: $NUM_{f,j} = \frac{ANUM_{f,j}-1}{N_f-1} * 4 + 1$. This variable is an afine transformation of $ANUM_{f,j}$. We subtract 1 from both the numerator and denominator of the ratio in the formula so that this ratio ranges from 0 to 1. By multiplying this ratio by 4 and adding 1, the resulting measure ranges from 1 to 5 in all business functions of all establishments, hence allowing for comparability.

$MAX_{f,j}$: $MAX_{f,j}$ denotes the sophistication level of the most-sophisticated technology used by the establishment in the BF. It is calculated using the formula: $MAX_{f,j} = \frac{r_{f,j}^{MAX}-1}{R_f-1} * 4 + 1$, where $r_{f,j}^{MAX}$ is the absolute rank of the most sophisticated technology used in a BF, and R_f is the rank of the most sophisticated technology **possible** in the BF f. For most BFs, R_f equals N_f.[45] However, there are a few exceptions where the number of technologies in the grid for ta BF does not equal the rank of possibly most sophistication technology. These exceptions are discussed in Appendix B.1.1. We apply the same afine transformation to the relative ranks so that the range of $MAX_{f,j}$ is $[1, 5]$.

$MOST_{f,j}$: $MOST_{f,j}$ denotes the sophistication level of the most widely-used technology in a particular BF by the establishment. It is calculated using the formula: $MOST_{f,j} =$

[44] For instance in the example in Figure B.1, 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.

[45] For example, if a respondent answered "Yes" to "Breed substitution", "Inbreeding", and "Artificial Insemination", and "No" for all other technologies, $r_{f,j}^{MAX}$ would be 3 for that establishment for this BF. Here $R_f = N_f = 4$, so $MAX_{f,j}$ would be 3.67.

$\frac{r_{f,j}^{MOST}-1}{R_f-1} * 4 + 1$, where $r_{f,j}^{MOST}$ is the absolute rank of the most widely-used technology in a BF, and R_f is the rank of the most sophisticated technology possible in BF f.[46] The relative sophistication ranking is subject to an afine transformation so that the range of $MOST_{f,j}$ is $[1, 5]$.

After constructing the measures at the establishment-BF level, we construct establishment-level measures by averaging them across all the BFs of each establishment. In particular, NUM_j, MAX_j, and $MOST_j$ are 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.5, 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.5). 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.2 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

[46] For example, if an establishment that uses "Breed subsbtitution", "Inbreeding" and "Artifical insem.", but uses "Inbreeding" most often, $r_{f,j}^{MOST}$ would be 2. Consequently, $MOST_{f,j}$ would be 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.2: 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.2 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.2: 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.10). 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.

C Additional Figures and Tables

C.1 Technological Sophistication

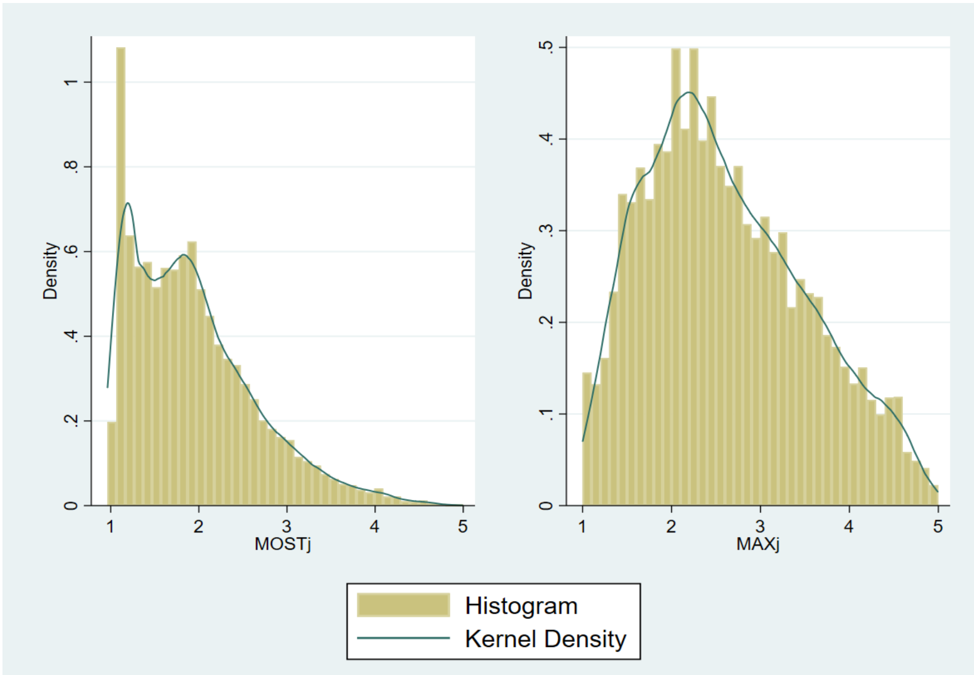


Figure C.3: Distribution of MAX_j and $MOST_j$ across establishments

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.1: 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.2: Average level of technology measures

Business Function	$ANUM_{f,j}$	$NUM_{f,j}$	$MAX_{f,j}$	$MOST_{f,j}$	N_f	Business Function	$ANUM_{f,j}$	$NUM_{f,j}$	$MAX_{f,j}$	$MOST_{f,j}$	N_f
General Business Functions						*Automotive*					
Business Administration	2.1	2.1	3.0	2.5	5	Vehicle Assembly	2.5	2.2	2.9	1.2	6
Production Planning	1.9	1.9	2.7	2.2	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	2.3	1.7	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	3.2	2.1	3
Payment	2.9	2.2	3.5	2.7	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	3.0	2.6	4
Irrigation	1.9	1.7	2.7	2.4	6	Weighing scale	1.9	2.2	4.1	3.7	4
Weeding	1.9	1.9	2.4	2.0	5	Mixing/Compounding	1.8	1.8	2.9	2.4	5
Harvesting	1.9	1.9	2.7	2.1	5	Encapsulation	1.8	2.1	3.8	3.1	4
Storage	1.8	1.8	2.7	2.3	5	Quality control	1.6	2.3	3.3	2.7	3
Packaging	1.6	1.8	2.1	1.7	4	Packaging	1.5	2.0	3.3	2.8	3
						Fabrication	2.0	1.8	2.3	1.9	6
Livestock						*Wholesale and Retail*					
Breeding	2.0	2.4	3.0	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	3.0	5	Merchandising	1.6	1.8	2.1	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	2.1	2.4	3.5	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	3.5	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	2.6	1.9	3	Planning	1.6	1.8	2.1	1.6	4
Cutting	2.0	2.0	2.4	1.9	5	Plan execution	1.9	1.9	2.2	1.5	5
Sewing	2.1	2.1	2.8	2.4	5	Monitoring	1.9	1.9	2.4	1.6	5
Finishing	1.7	1.7	2.1	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	2.4	2.0	3	Infrastructure and Machines *	5.5	2.8	2.8	2.8	11
Cutting	2.2	2.2	2.6	2.1	5	Appointment and Scheduling	2.0	2.3	2.6	1.7	4
Sewing	2.2	2.2	2.5	2.1	5	Patient records management	1.9	2.2	3.0	2.4	4
Finishing	1.6	1.6	1.8	1.5	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 reports the weighted average of the variables listed in the top row for each business function. The weights used are the sampling weights.

Table C.3: 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 in each business function. The percentages are calculated using sampling weights.

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.