Internet Adoption, Gender and Productivity in Informal Firms in India

Jayati Chatterjee[1] and Subash S[2] and Rajesh Raj S.N. [3]

[1]Corresponding author. PhD Research Scholar, Department of Humanities and Social Sciences, IIT Madras, India. Jchatterjee95@gmail.com
And
[2]Professor, Department of Humanities and Social Sciences, IIT Madras, India.
subash@iitm.ac.in
And
[3]Associate Professor, Department of Economics, Sikkim University.

Abstract
This study examines the impact of internet adoption on firm productivity in the unorganised sector in India through a gender lens. Prior research has highlighted persistent productivity differences between male- and female-led enterprises, but evidence on how internet adoption influences these differences remains limited. Using data from the NSSO surveys spanning 2010-11 to 2023-24, we observe that productivity has increased by an average of 28 per cent among informal sector firms that have adopted the internet. Our findings also confirm the persistence of gender-based productivity gaps: female-owned firms, on average, exhibit substantially lower productivity than male-owned firms by almost 48 per cent. However, internet adoption is positively associated with productivity among enterprises owned by women. Internet adoption yields an estimated 23 per cent productivity gain for the female-owned enterprises that adopt the internet, thereby narrowing the gap relative to non-adopters. These results highlight the importance of digital technology in bridging productivity gaps in the informal sector, offering valuable insights for policymakers aiming to foster inclusive digital growth.

Keywords:

informal sector, small firms, digitalisation, gender

1. Introduction

A substantial body of literature highlights a well-established link between ICT adoption and productivity growth across countries. Studies in developed countries consistently report productivity gains driven by improvements in ICT tools and IT-related industries. This trend has been documented in the United States (Jorgenson & Stiroh, 2000; Oliner & Sichel, 2000; Timmer & van Ark, 2005), the Netherlands (Borowiecki et al., 2021), Italy (Hall et al., 2013; Nucci et al., 2023), and other European economies (Gal et al., 2019). OECD reports (2020, 2021) and other studies (Gette et al., 2022) provide additional evidence. While most research focuses on developed countries, a strand of studies suggests that there are significant benefits to ICT adoption in developing countries as well (Fagerberg, 1994; Steinmueller, 2001). But the returns to productivity growth from ICT usage vary across developing and less developed countries (Dedrick et al., 2013).

The recent period has witnessed a surge in studies exploring the impact of specific ICTs, like broadband internet adoption (Grimes et al., 2012), mobile technology (Bertschek & Nibel, 2016), and internet usage (Clarke et al., 2015; Kretschmer, 2013), owing to the availability of firm-level data. These studies demonstrate a positive impact of such technologies, including enhanced customer service (Stansfield & Grant, 2003) and improved transactions (Vickery et al., 2004). However, available evidence is largely confined to the formal sector, with limited attention devoted to understanding ICT adoption among informal enterprises in developing economies, where these firms face considerable barriers to growth and technology adoption. Although the empirical relevance of digital adoption and performance is well established, there is a notable lack of studies examining this association in the context of the informal sector in developing countries.

The NCEUS (National Commission for Enterprises in the Unorganised Sector) defines an informal sector firm in India as a private, unincorporated enterprise with fewer than ten workers. In the case of firms that operate without power, the definition of informal firms includes those with twenty or fewer workers. Informal sector firms account for a substantial share of economic activity in developing countries, yet they often operate with outdated technology and encounter multiple barriers that hinder their productivity. Among them, female-led businesses encounter more challenges than male-owned enterprises (Rodriguez et al, 2023). Existing evidence shows that firms led by women consistently underperform compared to male-owned enterprises (Gang et al., 2022; Islam & Amin, 2023) due to their organisational structure, small size, and concentration in low-profit sectors (Klapper & Parker, 2011; World Bank, 2009). Most of these firms operate from household premises and rely solely on family labour, further constraining their ability to grow and transition to larger firms (Sabarwal & Terrel, 2008; Bardasi et al., 2011; Chaudhuri et al., 2020). Although gender productivity differentials have long been a subject of inquiry, researchers have only recently begun to examine these gaps in the context of the informal sector (Gang et al., 2022; Islam & Amin, 2023).

A tangential issue is the well-documented gender divide in technology adoption, which further hampers the performance of female-owned enterprises. Prior studies consistently demonstrate that women are less likely than men to adopt and use digital technologies due to societal norms, limited access to technical training, and low confidence in technology usage (Ahuja & Thatcher, 2005; Venkatesh & Morris, 2000; Maruping & Magni, 2012; Venkatesh et al., 2000; Dong & Zhang, 2011). However, the COVID-19 pandemic triggered a significant shift: evidence reveals that women entrepreneurs are increasingly using digital platforms to sustain

their businesses despite limited investments in software and hardware (Torres et al., 2021). This accelerated adoption of digital technologies, particularly internet usage, demonstrates the potential to reduce longstanding gender gaps in firm productivity.

Against this backdrop, this study addresses three critical questions. First, does internet adoption lead to higher productivity outcomes for informal sector firms in India? Second, does ICT adoption help mitigate productivity differences between male- and female-owned enterprises? Third, how are firm-level and owner-level attributes related to productivity outcomes? We examine these questions using a large, nationally representative dataset on informal sector firms in India, covering the period from 2010 to 2023.

This study contributes to the expanding body of literature on the relationship between internet adoption and firm-level productivity in the informal sector, with a focus on female-owned enterprises. The informal sector contributed approximately 45 per cent of the GDP in 2022-23 and employed over 110 million people (ASUSE, 2022-23). Informal sector firms are generally technologically less advanced (Mukherjee, 2018), and women have historically been underrepresented in entrepreneurship.[1] However, the share of female-led enterprises has grown significantly from 17.4 per cent in 2010 to 26.2 per cent in 2024.

Alongside this growth in female entrepreneurship, the Indian economy has witnessed a significant digital boom. According to the State of India's Digital Economy Report (2024), India now ranks third globally in digitalisation, and nearly 26.7 per cent of informal enterprises report using the internet (ASUSE Survey 2023-24). These trends, combined with the growth of small enterprises and the rising participation of women in entrepreneurship, make India an ideal setting for this study.

Recent policy initiatives have strengthened these trends by promoting digital adoption, boosting access to credit and encouraging new entrepreneurs. Credit-related policies, such as the Credit Guarantee Trust Fund for Micro & Small Enterprises (CGTMSE) and the Credit-Linked Capital Subsidy Scheme (CLCSS), aim to facilitate operations and technology adoption in micro, small and medium enterprises (MSMEs). Similarly, the Digital India scheme, initiated in 2015, seeks to expand internet availability and access nationwide. These policies may have contributed to the increasing number of entrepreneurs during the last decade (PIB, 2025).

Against this backdrop, this study makes two contributions. First, it examines the impact of internet adoption on productivity among informal firms using large-scale firm-level data covering 14 years. Second, while existing studies have primarily focused on the productivity differences between male- and female-owned businesses, this study investigates whether internet adoption can mitigate these gender-based disparities.

The main findings of this study reveal a significant productivity gap between ICT adopters and non-adopters, even after controlling for other firm-level factors. Labour productivity is 28 per cent higher for internet-adopting firms. Female-owned enterprises, on average, exhibit almost 48 per cent lower productivity than male-owned enterprises. More importantly, the adoption of digital technology significantly improves the productivity of female-led enterprises, helping to narrow the performance gap between men and women. Female-owned firms that use the internet are 23 per cent more productive than others. These results demonstrate that technological integration in small enterprises can significantly enhance operational efficiency. This finding is consistent with existing studies on gender differences in firm performance, although the literature on this topic is sparse. Furthermore, our findings suggest that although female entrepreneurs may face greater challenges in adopting technology,

[1] The total number of such enterprises rose from 57.7 million in 2011 to 67 million in 2024 (ASUSE 2023-24).

leveraging digital tools can help bridge the productivity gap and enhance their resilience, particularly during the post-pandemic period. Due to the paucity of firm-level information, we are unable to fully understand internet usage among small firms; however, our results are robust in indicating a strong relationship between internet adoption and productivity outcomes.

The rest of the paper is organised as follows: Section 2 outlines a brief literature review. Section 3 discusses the data and provides descriptive evidence of the unorganised enterprises in the Indian economy. Section 4 presents the methodology used to understand the impact of ICT adoption on labour productivity, and Section 5 addresses the productivity differentials across gender and how ICT adoption could reduce them through regression analysis. Section 6 concludes the paper.

2. Related Literature

2.1 ICT and Productivity

Information and communication technologies (ICT) have played a crucial role in shaping modern economies (Hall et al., 2013; Choi & Hoon Yi, 2009). Macro-level studies have highlighted that the internet has been instrumental in promoting economic growth and development during the last few decades. Accelerated reliance on ICT has yielded benefits in both developed and developing nations, with increased use of digital technologies (Yousefi, 2011; Hitt & Brynjolfsson, 1996). While developed economies have widely adopted these technologies, many developing nations are beginning to do so through targeted policy measures (Dedrick et al., 2013). In developed countries, technological innovations associated with the ICT revolution have had wide-ranging spillover effects in various sectors. The ICT revolution has resulted in wide-ranging spillover effects across various sectors and outcomes (Cardona et al., 2013), extending beyond firms to include healthcare (Shao et al., 2022) and education (Fernández-Gutiérrez et al., 2020).

In the context of industrial development, the adoption of ICT has been crucial for firm growth and productivity (Niebel, 2018; Stiroh, 2002). For example, Stiroh (2002) observe that IT industries investing in ICT experience positive spillovers through higher labour productivity. Echoing these results, Brynjolfsson and Hitt (2003) demonstrate that technology adoption made a significant contribution to output growth and productivity in the US between 1987 and 1994. Bloom et al. (2010) highlight cross-country differences and demonstrate that firms with ICT are more likely to grow faster and less likely to exit the market in the UK and other European countries. Firm-level studies further identify specific ICT and other firm-level characteristics that explain productivity gains in these countries (Bertschek & Niebel, 2016; Borowiecki et al., 2021; Gal et al., 2019). Recent studies, such as Cette et al. (2022), Nucci et al. (2023), and Abramova and Grishchenko (2020), report a significant positive impact of investments in digital technology on productivity. The OECD (2021) notes that low-productivity firms tend to benefit more from investments in technology.

In contrast, the association between technology adoption and productivity growth in developing countries is relatively recent, mainly due to the historical lack of firm-level data for small firms. However, more recently, the availability of firm-level data has enabled researchers to explore the impact of higher ICT investments on productivity (Aboal et al., 2017; Mushtaq et al., 2022; Afolayan, 2016; Dutta et al., 2023; Mishrif & Khan, 2023; OECD, 2024). In the context of India, Erumban and Das (2016) use sectoral data to show that ICT positively affects economic growth, and most of this effect is concentrated in the services sector. Gupta and Kumar (2018) report that ICT adoption has benefited larger firms in India more than small and medium-sized firms.

2.2 Gender gap in performance

Our study is related to existing research on the gender gap in firm performance (Delecourt & Ng, 2021; Sabarwal & Terrel, 2008; Chaudhuri et al., 2020; Gang et al., 2022; Islam & Amin, 2023). These studies highlight both observable and unobservable factors driving these differences. Some studies show that gender gaps in earnings among firms can be explained by observable factors such as low inventory (Delecourt and Ng), limited cash availability (Mel et al., 2008) and smaller scale of operations (Khalife & Chalouh, 2013; Sabarwal & Terrel, 2008; Bardasi et al., 2011). At the same time, some studies find that gender gaps in performance persist even after controlling for these factors (Robb & Watson, 2004). In some cases, female-owned enterprises outperform male-owned enterprises, which is often attributed to higher returns to scale associated with small firm size (Watson & Robinson, 2003; Sabarwal & Terrel, 2008). Other studies indicate that female enterprises are predominantly concentrated in low-capital-intensive sectors (Klapper & Parker, 2011) or operate from household premises, which may result in reduced profits (World Bank, 2009).

Household responsibilities and time allocation further influence these disparities. Studies have found that women often spend less time on their business activities compared to men, which contributes to lower firm performance (Nix et al., 2015; Robb, 2008; Hundley, 2001). The unequal distribution of household responsibilities is a critical factor driving this gap (Bruhn et al., 2009; Fafchamps et al., 2014; Delecourt & Fitzpatrick, 2019; Cohoon et al., 2010; Klapper & Parker, 2011).

The lack of motivation, aversion to competition, and risk preferences also help to explain the gender gaps in firm performance (Coohoon et al., 2010; Klapper & Parker, 2011; Gneezy et al., 2003) as well asexpansion strategies (Rosa et al., 1996). Women business owners are also more protective of their intellectual capital and rely heavily on networking (Cohoon et al, 2010). Furthermore, studies suggest that women are more likely to perceive the business environment as hostile, which can discourage growth initiatives such as adopting new technologies (Expisito et. al., 2023) or introducing innovations (Na & Shin, 2019).

However, evidence is not uniform. For the United States, Robb and Watson (2004) find no differences in closure rates and in returns to assets between men and women. Another study by Watson and Robinson (2003) shows that women outperform men. Hardy and Kagy (2018) similarly point out that other factors may be at play, as they did not find performance differentials explained by firm-level and owner characteristics.

3. Data Source and Variable Description

3.1 Data

This study draws on five rounds of the National Sample Survey Office (NSSO) surveys of unincorporated non-agricultural enterprises. We utilise two rounds from the NSS Non-Agricultural Enterprise Survey, specifically the 67th (2010-11) and the 73rd (2015-16) rounds, and three rounds from the Annual Survey of Unincorporated Sector Enterprises (ASUSE) for 2021-22, 2022-23, and 2023-24.[2] The 2010-11 and 2015-16 rounds follow the earlier quinquennial pattern and the latest three rounds form part of the Annual Survey. Until 2021, the NSSO conducted periodic surveys of enterprises in the non-agricultural sector. In response to the demand for more frequent data, the ASUSE was introduced in 2021 to provide annual

[2] For the sake of simplicity, we refer the survey conducted during 2010 - 2011 as 2010, 2015 -16 as 2015 and so on.

estimates of enterprise activity. Both surveys focus exclusively on non-agricultural enterprises.[3]

The NSSO defines unincorporated enterprises as those not registered under the Companies Act of 1956 or 2013 and hence not treated as corporate entities. These surveys cover manufacturing enterprises not registered under the Factories Act of 1948 as well as those engaged in trade and other services. A small subset of enterprises with partial registration under regulations, such as Section 85 of the Factories Act or state-level Shops and Establishments Act is also included, though these units remain outside the corporate and factory-registered segments. Thus, the dataset pertains entirely to informal or unorganised enterprises and does not cover formal sector firms.

The surveys systematically collect operational, financial, and structural information at the establishment level for manufacturing and service entities. Since all five survey rounds are cross-sectional, we create a pooled dataset for empirical analysis. The number of firms surveyed varies across years: 334,474 in 2010-11, 290,113 in 2015-16**,** 4,16,269 in 2021-22, 4,58,938 in 2022-23 and 4,98,024 in 2023-24. Altogether, this amounts to nearly 2 million firm observations, which reduce to 1.6 million after data cleaning.

.

As part of the filtering process, we exclude non-profit enterprises and limit the analysis to firms located in 17 major states that account for the largest share of enterprises.[4] The final sample also excludes partnerships and self-help groups and retains only single-owner firms. We further restrict the sample to firms that are less than 100 years old and employ fewer than 20 workers.[5]. The total sample included in our analysis after cleaning is 264,494 in 2010-11, 221,192 in 2015-16, 337,889 in 2021-22, 365,634 in 2022-23, and 398,390 in 2023-24.

3.2 Variables

We explain below the construction of the variables used in this study:

Dependent Variable

Our primary measure of firm performance is labour productivity. Labour productivity is computed as the logarithmic value of the ratio of gross value added[6] to the total number of workers, following previous studies (Rabbani & Raj, 2024). Gross value added is calculated as output minus intermediate inputs, both of which are provided in the dataset. The total number of workers includes both formal and informal workers, as well as the owner.[7].

Independent Variables

---

[3] The non-agricultural enterprises are engaged in the activities of manufacturing, construction, trading and repair services, hotels and restaurants, transport storage and communications, financial intermediation, real estate, renting and business activities, education, health and social work, other community, social and personal service activities (excluding domestic services).

[4] The states include Punjab, Haryana, Rajasthan, Uttar Pradesh, Bihar, Assam, West Bengal, Jharkhand, Odisha, Chhattisgarh, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Telangana, Karnataka, Kerala, Tamil Nadu. The states of Andhra Pradesh and Telangana have been considered as one. These states are selected out of the major states reported in the ASUSE Report 2023-24, whose percentage in the total share of enterprises in the country was more than 1.6 %.

[5] Firms more than 20 workers are generally considered small according to World Bank, hence we limit ourselves with only 20 workers.

[6] The gross value added is inflation-adjusted and winsorized at 1 and 99 percentiles.

[7] We consider only permanent workers since the labour productivity may be biased if we include the temporary workers.

Internet adoption: We measure digital technology adoption using a proxy for internet adoption. The variable is based on the question of whether the enterprise uses the internet (Yes or No). We assign the value one if the firm reports internet use and zero otherwise. Since the literature indicates that ICT adoption has positive spillovers on firm performance, we expect internet adoption to have a positive association with labour productivity.

Gender: The variable refers to the gender of the individual who owns and manages the enterprise. We create a binary variable that takes the value of one if the firm is owned and managed by a female and zero otherwise. Based on previous studies, we expect the labour productivity of female-owned firms to be lower than the male-owned enterprises

Other Controls

We control for the social group affiliation of the firm owner, which is defined by the survey using four caste categories, but we used here as athree category variable: OBC (Other Backward Classes), SC/ST (Scheduled Castes and Scheduled Tribes, merged into a single category due to small sample sizes) and the general category, which serves as the reference group.[8]

Other firm characteristics include the location of the establishment and firm size. Location is represented using two binary variables. The first indicates whether the enterprise operates within or outside the household, taking the value 1 if it operates outside and 0 if it operates within the household. Units classified as fixed with temporary structure, fixed with permanent structure, fixed with no structure, mobile market, and street vendor are considered operating outside the household. The second location variable distinguishes rural and urban firms, and is coded 0 for rural firms and 1 for urban firms.

For firm size, we use a dummy variable indicating whether the firm is an own-account enterprise, taking the value 0 for own-account enterprises and 1 for hired worker establishments. The age of the enterprise is calculated by subtracting the year of initial operation from the survey year. To compute the log of age, we add one to the age value before taking the logarithm. We control for the financial condition of firms using a binary variable that takes the value one if the firm reports any financial constraints and zero otherwise. We also include state and two-digit NIC industry-year fixed effects.

Table 1[9] provides the summary statistics for all variables used in the analysis. Productivity is higher among internet users compared to non-users. Female-owned firms, on average, have lower productivity than male-owned firms. These observations are consistent with trends in firm ownership and internet adoption over time. As shown in Figure 1, the share of female-owned enterprises increased from 19 per cent in 2010 to 28 per cent in 2023. Over the same period, the share of internet adopters also increased substantially among both male- and female-owned enterprises (Figure 2). The graphical figures are computed using sample weights; therefore, the mean may differ from that in the descriptive statistics in Table 1.

Table 1: Descriptive Statistics

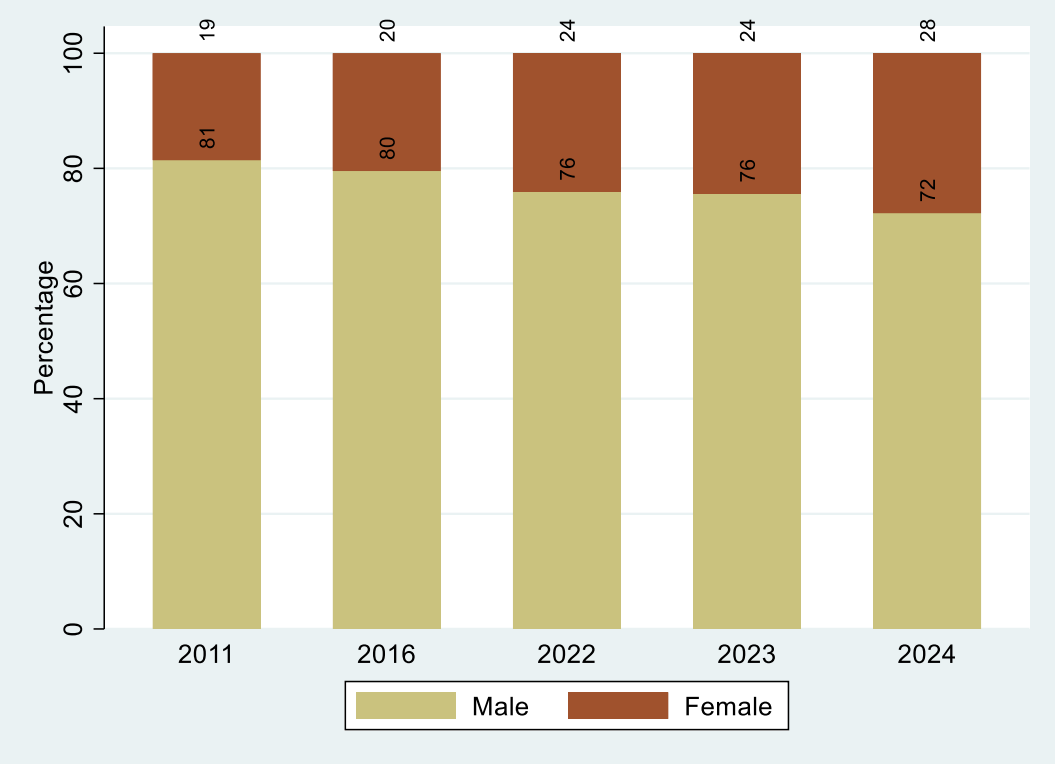
Variable	Obs	Mean	Std. Dev.	Min	Max
Log of Productivity	1568263	11.14	0.88	1.46	18.93
Log of productivity for non-	1287622	11.04	0.88	1.46	16.32

[8] This classification follows the standard NSSO categorisation and is based on constitutional provisions and official government notifications.

[9] The summary statistics represent the actual number of observations in the sample. While the graphs represent the weighted values of the sample.

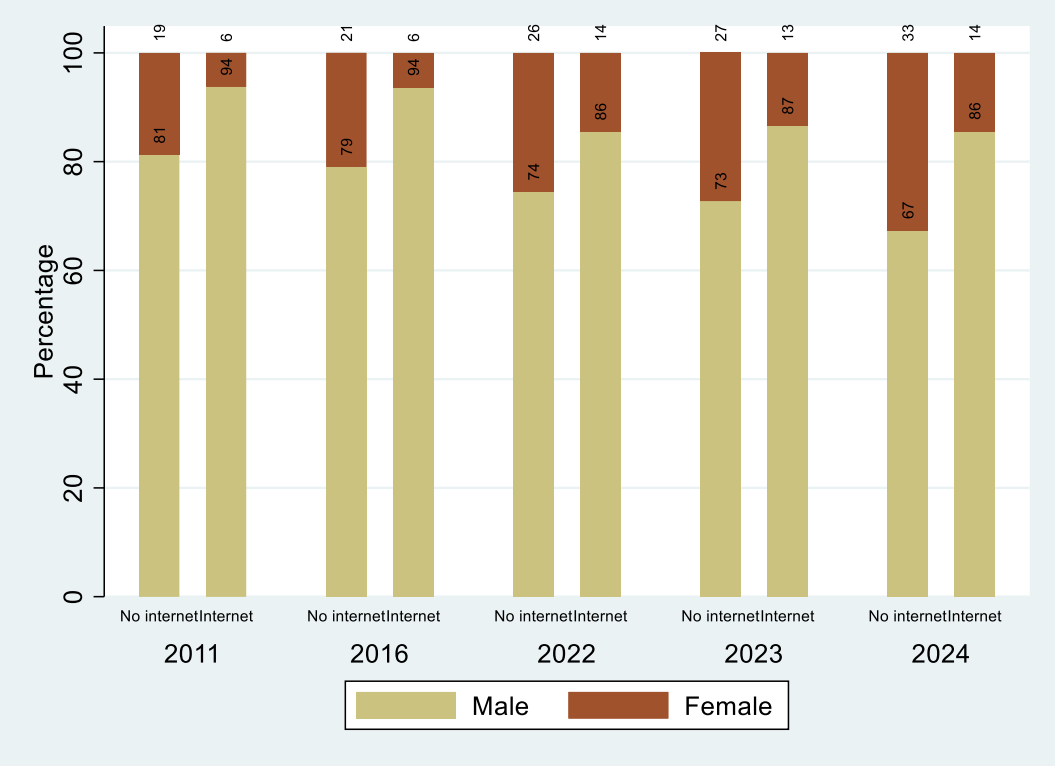
Internet users					
Log of productivity for internet users	280641	11.61	0.72	2.15	18.93
Log of productivity for male-owned firms	1270449	11.31	0.76	2.06	18.93
Log of productivity for female-owned firms	297814	10.41	0.98	1.46	15.18
Internet	1568263	0.18	0.38	0	1
Female	1568263	0.19	0.39	0	1
General/others	1568263	0.31	0.46	0	1
OBC	1568263	0.54	0.50	0	1
SC and ST	1568263	0.15	0.36	0	1
Proportion of enterprises outside the household	1568263	0.65	0.48	0	1
Proportion of urban enterprises	1568263	0.47	0.50	0	1
Proportion of Hired worker establishments	1568263	1.26	0.44	1	2
Log of age	1568263	11.41	8.60	2	100
Proportion of enterprises with financial obstacles	1568263	0.07	0.25	0	1

Figure 1: Share of enterprises by the gender of the owner (%)



Source: Authors computed from NSSO and ASUSE survey

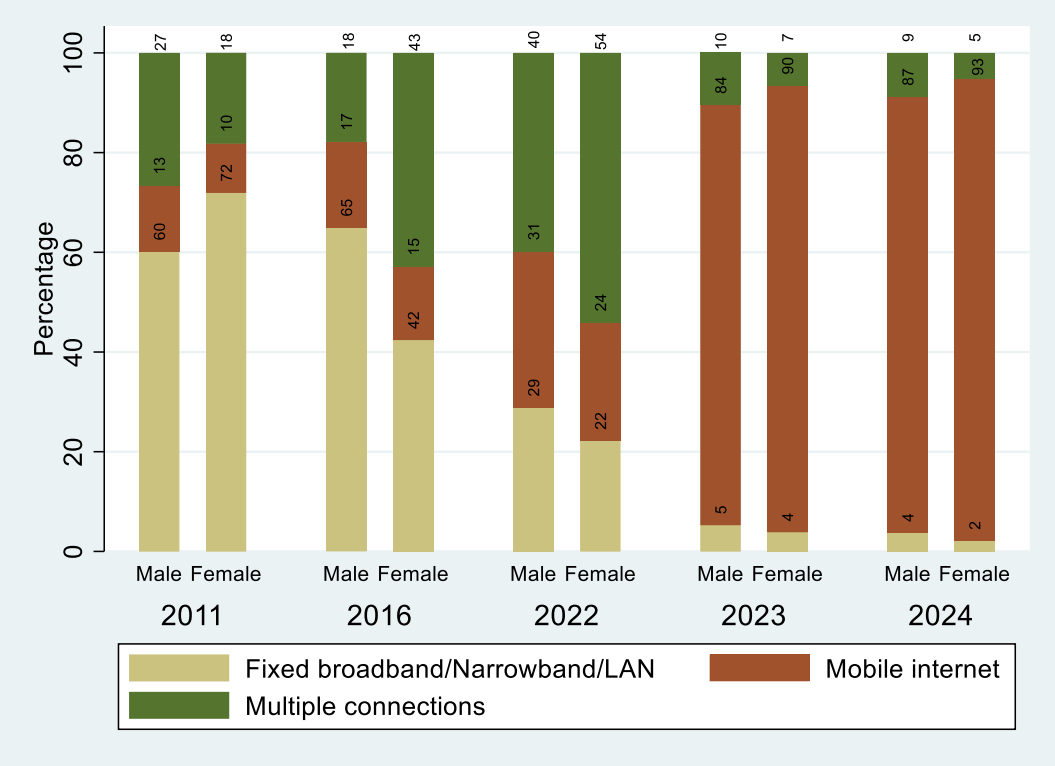
Figure 2: Internet adoption by Ownership (%)



Source: Authors computed from NSSO and ASUSE survey

Our dataset provides information on various types of internet connections: (i) non-mobile types, which include narrowband, broadband, and Local Area Network (LAN)(ii) mobile internet, and (iii) multiple connections combining mobile and fixed-line. During the study period, we observed a marked transition: non-mobile connections have declined sharply, accounting for only 2–4% of firms using internet in recent years. This marks a clear shift from earlier periods when broadband and LAN connections were dominant (Figure 3), as mobile internet has rapidly displaced fixed-line adoption.

Figure 3: Type of internet connection by ownership



Source: Authors computed from NSSO and ASUSE survey

Mobile internet has emerged as the dominant form of internet connectivity, particularly over the past two years (Sahoo et al., 2021; Zhang & Hu, 2024). Several factors drive this shift. First, mobile internet is relatively affordable, making it accessible even to small, resource-constrained enterprises. Second, it requires minimal infrastructure, unlike broadband or LAN setups that often involve installation charges, hardware requirements, and technical support. In contrast, mobile data can be accessed with basic smartphones, which are now widely used[10]. Moreover, mobile internet is less vulnerable to disruptions from external factors such as power outages or adverse weather, which disproportionately affect fixed-line connections. This resilience is especially valuable in regions with inconsistent electricity, where mobile connectivity helps sustain business operations. Critically, this transition extends beyond technological change: it carries gendered significance. Given that female-owned enterprises typically operate at smaller scales with limited resources, affordable mobile internet likely plays a pivotal role in enabling their digital economic participation.

4. Methodology

4.1 OLS and 2SLS Regression

To understand the nexus between digitalisation and firm performance, we begin by conducting an exploratory analysis of our data using a kernel density plot. Subsequently, we employ an OLS estimation followed by an instrumental variable (IV) approach to investigate the determinants of firm performance[11]. We estimate the following OLS model:

$$lnLP_{ijst} = \alpha_1 + \alpha_2 \times Gender_{ijst} + \alpha_3 \times Internet_{ijst} + \alpha_4 \times Gender_{ijst} \times Internet_{ijst} +$$

[10] According to Waghmare (2024), in 2023, 85 % of Indians reported having a smartphone, and only in 2023, eight out of ten homes were connected to the internet.

[11] All regression uses the sample weights provided in the survey.

$\alpha_5 \times X_{ijst} + \alpha_6 \times Z + \mu_{ijst}$

The dependent variable ($lnLP_{ijst}$) is the log of labour productivity for the *i*-th firm with two-digit NIC code *j*, located in state *s* at time *t*. The variable *Gender* is a dummy variable indicating whether the firm is female-owned. The variable $Internet_{ijst}$ is a dummy variable denoting internet adoption, which takes the value 1 for adopters and 0 for non-adopters. The interaction term $Gender_{ijst} \times Internet_{ijst}$captures whether internet adoption in female-owned firms is associated with higher or lower productivity compared to female-owned firms without internet.

The vector *X* includes a host of other firm controls defined earlier, including age of the firm, social group of the owner, location of the enterprise represented by two binary varibales whether the firm operates within household or outside the household and whether it is located in a rural or urban area, type of enterprise (hired worker establishment or own account enterprise), and type of industry (manufacturing or non-manufacturing). The vector Z captures the industry-year and state fixed effects.

A key concern with our baseline model is that internet adoption may be endogenous. Firms with higher productivity might be more likely to adopt the internet, thereby creating a reverse causality problem. Further, unobserved factors such as managerial skill and owner motivation may also influence both productivity and ICT adoption, leading to omitted variable bias.

To address this, we employ a two-stage least squares (2SLS) approach, which mitigates the endogeneity bias arising from possible reverse causality between ICT adoption and productivity.

The instrumental variable (IV) approach involves two stages. In the first stage, we regress ICT adoption on the instrumental variable, along with firm-specific characteristics and ownership variables, while controlling for state, industry, and year effects. The model is specified as follows:

$$Internet_{ijst} = \beta_0 + \beta_1 Gender_{ijst} + \beta_2 X_{ijt} + \beta_3 IV_{ijst} + \phi_{ijst} \quad (2)$$

where $Internet_{ijst}$ is a dummy variable for internet adoption, IV_{ijst} denotes the instrument, and *X* represents firm-level controls as in Equation (1).

We use the share of ICT adopters within the same industry, region, and year as an instrument for internet adoption. This measure, referred to in the literature as the "epidemic effect", captures the tendency of a firm to adopt ICT when many neighbouring firms in the same industry and region have already adopted it, thereby capturing learning from peers and local spillovers (instrument relevance). We consider the instrument exogenous after controlling for industry-year , and statefixed effects and argue that it satisfies the necessary exclusion restriction as it influences firm productivity only through ICT adoption. We test the validity and suitability of the instrument in our estimations. Following Haller and Siedschlag (2011), we estimate these epidemic effect as:

$$epid_{indstatyr} = \frac{X^{\propto}_{jdt}}{X_{jdt}}$$

where $X^{\propto}_{jdt}$ is the number of adopters in industry j, district d, and time *t*, and X_{jdt} is the total number of firms in that group. We use the district as a region, as most of these states are quite large, and geographical proximity is an important determinant of internet usage due to peer

effects. Thus, $epid_{indistyr}$measures the proportion of adopters within each industry-region-year group, which captures the intensity of diffusion or peer effects in the local industry environment, reflecting the extent to which widespread adoption is observed among potential adopters in the same context (industry-region-year).

A higher value of $epid_{indstatyr}$ indicates that a larger share of firms in a given industry and location have adopted the innovation, which may generate spillover effects through learning, imitation, or competitive pressure. In empirical models, this variable test whether firms are more likely to adopt when adoption rate among their peers is higher, reflecting social learning.

Our second stage regression takes the following form:

$$lnLP_{ijst} = \gamma_1 + \gamma_2 Gender_{ijst} + \gamma_3 (\widehat{Internet})_{ijst} + \gamma_4 Internet_{ijst} \times Gender_{ijst} + \gamma_5 X_{ijst} + \alpha_6 Z_{ijst} + \epsilon_{ijst} \qquad (3)$$

The first stage regression in equation 2 yields predicted values of ICT adoption, $\widehat{Internet}_{ijst}$, which are included in the second-stage regression in Equation 3. The coefficient of the gender dummy captures the productivity gap between male-owned and female-owned firms, whereas the coefficient of the interaction term between gender and predicted ICT adoption indicates whether internet adoption affects this gap.

4.2 PSM Approach

Our primary identification strategy for estimating the causal effect of internet adoption on firm productivity relies on an IV approach. To ensure robustness, we complement the IV analysis with the Propensity Score Matching (PSM) approach. While the IV approach addresses potential endogeneity arising from unobserved variables, PSM balances observed characteristics between treated and control groups. The consistency of results across these methods, each relying on different identifying assumptions, strengthens our findings and suggests that estimated effects are not driven by model-specific biases or unobserved heterogeneity.

To address selection on observables, we employ a set of semi-parametric estimators, including inverse probability weighting (IPW), augmented IPW (AIPW), and nearest-neighbour (NN) matching. Internet adoption is modelled using a set of firm-level covariates, which includes social group affiliation of the firm owner, age, location, sector, industry category, and firm size.

The IPW method corrects for confounding by reweighting the sample such that the distribution of covariates becomes independent of treatment status. Using a logit model, we estimate the propensity score for each firm, which indicates the probability of adopting the internet. Let *T* denote the treatment indicator (internet adoption) and X the set of covariates. Confounding occurs when $P(T|X) \neq P(T)$. To correct for this, we use inverse predicted probability weights, $\frac{1}{P(T|X)}$, and estimate the model. This reweighting approximates the conditions of a randomised experiment and allows us to assess the Average Treatment Effect (ATE) of internet adoption on productivity under the assumption of selection on observables. The IPW estimator does not rely on functional form assumptions for the outcome model, and the estimated treatment effect applies to the entire population. The ATE using IPW is given by:

$$ATE_{IPW} = \hat{T}_{IPW} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{T_i Y_i}{\hat{e}(X_i)} - \frac{(1-T_i)Y_i}{(1-\hat{e}(X_i))}\right]$$

where, Y_i is the log of labour productivity for firm *i*, and $\hat{e}(X_i)$= P(T=1|X_i) is the estimated propensity score.

X_i= Covariates

To mitigate possible misspecification in the propensity score model, we also employ the augmented IPW (AIPW) estimator, which combines the weighting mechanism of IPW with an outcome regression model for productivity based on the same covariates. This doubly robust approach produces consistent estimates if either the propensity score or the outcome model is correctly specified:

$$ATE_{AIPW} = \hat{T}_{AIPW} = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\widehat{u_1}(X_i) - \widehat{u_0}(X_i)\right) + \frac{T_i(Y_i - \widehat{u_1}(X_i))}{\hat{e}(X_i)} - \frac{(1-T_i)(Y_i - \widehat{u_0}(X_i))}{(1-\hat{e}(X_i))}\right]$$

Finally, as an additional robustness check, we implement PSM to construct matched samples of treated and untreated firms with similar observable characteristics. Firms are matched on propensity scores derived from the logit regression of treatment to ensure balanced covariate distributions between groups. The average treatment effect is estimated as:

$$ATE_{PSM} = \hat{T}_{PSM} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \frac{1}{M}\sum_{j\in J_M(i)} Y_j)$$

where, $J_M(i)$ denotes the set of M nearest neighbour matches for firm *i*

Together, these approaches, augmented by IPW and PSM, provide a robust framework to identify the causal effect of internet adoption on labour productivity while addressing selection bias and ensuring comparability between treated and control firms.

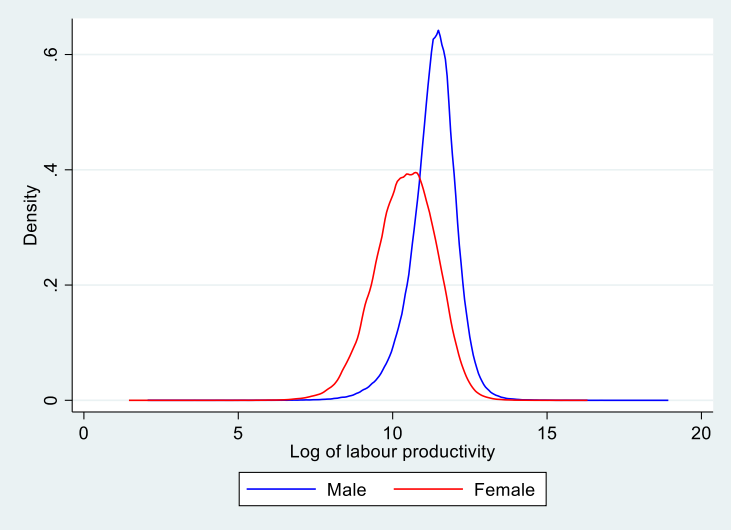
5. Results

5.1 Kernel density distribution

The descriptive statistics in Table 1 indicate that female-owned enterprises have lower average productivity than male-owned enterprises. Though the mean provides a general picture, it does not capture variation in productivity across distribution. To capture these gender differences in more detail, we present the kernel density plots in Figure 4.

Figure 4 displays the productivity distribution of male- and female-owned firms. At lower productivity levels, the distributions are similar, but the gap widens as we move toward the middle of the distribution. In the upper tail, the productivity distributions converge, which suggests that gender differences are more pronounced among firms in the lower and middle segments of productivity. The kernel density plots also reveal that women-led enterprises peak at lower productivity levels, while male-owned enterprises exhibit peak at higher productivity.

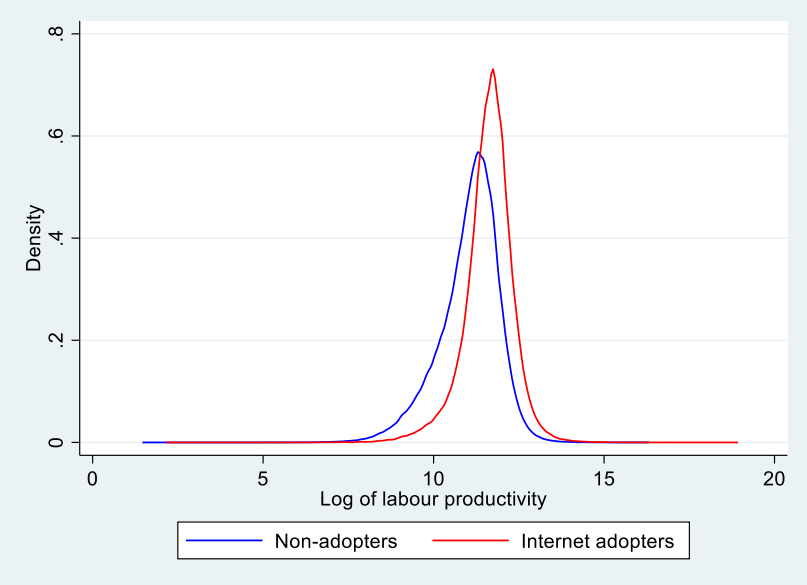
Figure 4: Kernel densities of productivity distribution by gender



Note: Kernel densities are computed for equispaced points using an Epanechnikov kernel
The X-axis measures the lnLP or the log of labour productivity, which is indicated as a variable in the figure.

Figure 5 presents kernel density plots of labour productivity by internet adoption. Firms adopting the internet generally exhibit higher productivity levels than non-adopters. The differences are most evident at lower productivity levels, whereas in the upper tail, which may include older and larger firms, the productivity gap diminishes. These patterns motivate us to analyse the impact of internet adoption on informal firms.

Figure 5: Kernel densities of productivity distribution across internet adopters



Note: Kernel densities are computed for equispaced points using an Epanechnikov kernel
The X-axis measures the lnLP or the log of labour productivity, which is indicated as a variable in the figure.

5.2 Results

This section presents the results of the empirical analysis, including OLS and IV estimation. Column (1) of Table 2 reports OLS results with only one explanatory variable, internet use, which is statistically significant at the 1% level. In the absence of controls and fixed effects,

the labour productivity of internet adopters is about 59 per cent higher than that of non-adopters.

In column 2, we introduce female ownership. The estimated effect of internet adoption falls to 50 per cent once the gender variable is included, suggesting that part of the initial difference is due to gender differences between firms that adopt and those that do not. The coefficient for female-owned firms is negative and significant (-0.75). After exponentiating the coefficient[12], the results suggest that female-owned enterprises have, on average, 53 per cent lower labour productivity than male-owned enterprises. This finding highlights a significant gender disparity in firm productivity, which is consistent with existing evidence on productivity gaps between male- and female-owned small firms (Bose, 2022; Gang et al., 2022; Islam & Amin, 2023).

Table 2: OLS Regression results

	Dependent Variable = Labour Productivity			
	(1)	(2)	(3)	(4)
Internet	0.465***	0.405***	0.375***	0.268***
	(0.021)	(0.016)	(0.018)	(0.017)
Gender		-0.75***	-0.77***	-0.658***
		(0.042)	(0.042)	(0.038)
Internet × Gender			0.191***	0.209***
			(0.04)	(0.033)
OBC				-0.084***
				(0.018)
SC/ST				-0.227***
				(0.035)
Premise: Outside the household				0.235***
				(0.03)
Sector: Urban				0.272***
				(0.027)
Hired worker establishment				0.131***
				(0.016)
Log age				0.07***
				(0.005)
Financial obstacle				-0.019
				(0.028)
Industry-Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Constant	10.932***	11.115***	11.121***	10.759***
	(0.003)	(0.01)	(0.01)	(0.022)
Observations	1581540	1581540	1581540	1581540
R-squared	0.315	0.381	0.381	0.431

Clustered standard error in parentheses

*** p<.01, ** p<.05, * p<.1

These results align with those of Boroweicki et al. (2018) and Gupta & Kumar (2018), who also find that digital adoption enhances the productivity of small firms. In Columns 3 and 4,

we introduce an interaction term between gender and internet adoption. The coefficient on the interaction term (0.209 in Col 4) indicates that female-owned firms that adopt the internet show a positive association with productivity. After exponentiating[13] this translates to an estimated 23 per cent higher labour productivity relative to non-adopting female-owned firms, beyond the individual effects of each variable. Thus, internet adoption reduces the labour productivity gap for female-owned firms even after controlling for firm-level characteristics. The overall gender log of the labour productivity gap stands at close to 0.658, and female-owned internet-adopting firms have a 0.209 higher log of labour productivity.

Regarding the other controls, we find that older firms are more productive than younger firms. Firms in urban areas, as well as hired worker establishments and those operating outside household premises, also exhibit higher productivity. Previous studies (Raj & Sen, 2016; Bigsten & Gebreeyesus, 2007; Liedholm & Mead, 1999) have associated larger and better-functioning firms with urbanisation. Social group affiliation also affects labour productivity. Using the general category (Others) as the base group, we find that the log of labour productivity for OBC-owned firms is about 8 per cent less and the SC/ST-owned firms are almost 22.7 per cent less.

After controlling for other firm-specific characteristics, the gap in labour productivity is 0.268 in log productivity. However, the interaction term indicates that internet adoption by female-owned firms is associated with approximately 23 per cent higher log productivity.

4.4 Instrument variable regression

To estimate the causal effect of internet adoption on firm productivity, we turn to the two-stage least squares (2SLS) results. The instrumental variable strategy and its motivation were discussed earlier. Table 3 presents the results for two specifications. Columns 1 and 2 use the share of ICT adopters in the same industry, region and year as the instrument. Table 3: The Impact of Internet Adoption on Firm Productivity (2SLS)

	FIRST INSTRUMENT $epid_{indstatyr}$	
	(1)	(2)
	INTERNET	LABOUR PRODUCTIVITY
	First Stage	Second Stage
Instrument: Share of ICT adopters[#] in the same industry-region-year	0.823***	
	(0.0118)	
Internet		0.272**
		(0.0800)
Gender	-0.112***	-0.657***
	(0.0201)	(0.0439)
Female Internet adopter	0.683***	0.205**
	(0.0181)	(0.0669)
OBC	-0.0185***	-0.0843***
	(0.00425)	(0.0189)
SC/ST	-0.0278***	-0.227***
	(0.00542)	(0.0356)
Premise: Outside the household	0.0179***	0.235***

[13] The exponential is calculated using $percentage\ change = \left(e^{\beta} - 1\right) * 100$, where β is the coefficient.

	(0.00308)	(0.0292)
Sector: Urban	0.0197***	0.272***
	(0.00262)	(0.0278)
HWE	0.117***	0.130***
	(0.00461)	(0.0206)
Log age	-0.0127***	0.0706***
	(0.00183)	(0.00407)
Financial obstacle	0.0140*	-0.0196
	(0.00593)	(0.0276)
Industry-Year FE	Yes	
State FE	Yes	
Constant	0.0337***	
	(0.00707)	
Observations	1581540	1581540
Kleibergen-Paap rk LM statistic	11.414	
	(0.007)	
Cragg-Donald Wald F statistic	3.9e+05	
Kleibergen-Paap rk Wald F statistic	4829.287	

State-level clustered standard errors in parentheses
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$
#ICT adopters refers to internet users

The first-stage results indicate that the instrument is positive and statistically significant, suggesting that higher local ICT penetration is strongly correlated with a firm's likelihood of adopting the internet. This finding aligns with prior evidence from Haller and Siedschlag (2011), Baptista (2000), and Keller (2002). The Kleibergen–Paap rk LM and the Kleibergen-Paap Wald F-statistic reveals that the model is not weakly identified. . Overall, these results suggest that the instrument is both statistically strong and suitable for predicting internet adoption (Table 3).

Table 3 (Columns 1 and 2) shows that the coefficient on internet adoption is positive and statistically significant. This implies that firms adopting the internet record 0.823 higher log labour productivity. The instrument also indicates that epidemic effects positively influence internet adoption: when more firms in the same industry, region, and year adopt the internet, a firm is more likely to adopt as well. . Female-led firms are significantly less productive, by almost 0.112 in terms of log of labour productivity. However, the gap in log of labour productivity narrows with internet use, as the variable of female-owned firms with internet compared to others have a positive and significant coefficient of 0.683 .is 0.683(0.2 The firm-level characteristics follows the OLS results in Table 2.

5.3 Propensity score matching

Firms that adopt the internet are likely to differ from non-adopters in both observed and unobserved dimensions**.** Following established practices in the digital adoption literature (Dutta & Sobel, 2016; Mallick & Yang, 2013), we employ propensity score matching (PSM) to construct a counterfactual framework that approximates experimental conditions. The matching techniques reduce such biases as discussed in the methodology section. We do not consider the cross-product term of internet and gender due to collinearity, and our focus remains on addressing the endogeneity concerns related to omitted variables. The PSM results provide an additional robustness check.

We first estimate the average treatment effect using inverse probability weighting (IPW)[14]. The IPW estimator reweights firms by the inverse of their estimated probability of treatment, constructing a pseudo-population where treatment assignment is independent of observed covariates. The ATE for the IPW shows that firms adopting internet exhibit 43.2% higher productivity after controlling for observable characteristics through inverse probability weighting.

Second, AIPW results are broadly consistent with the IPW estimates and the main IV findings, which supports the positive association between internet adoption and productivity. The additional robustness and efficiency that AIPW offers make it suitable for pooled cross-sections where firm characteristics vary across years and regions.

Table 4: Propensity score models

	(1) IPW	(2) AIPW	(3) PSM
No internet vs Internet ATE	0.451***	0.442***	0.287***
	(0.005)	(0.004)	(0.002)
N	1581544	1581544	554981

After matching, we assess covariate balance using standardised mean differences and variance ratios. The matched sample shows marked improvement in balance compared to the unmatched sample. While PSM does not account for unobserved heterogeneity and can introduce bias if matches are poor, it is a transparent and intuitive tool for comparing treated and control firms. The average treatment effect is positive and significant at the 1 per cent level, which supports our hypothesis that the internet adopters have about 0.287-0.45per cent higher productivity.

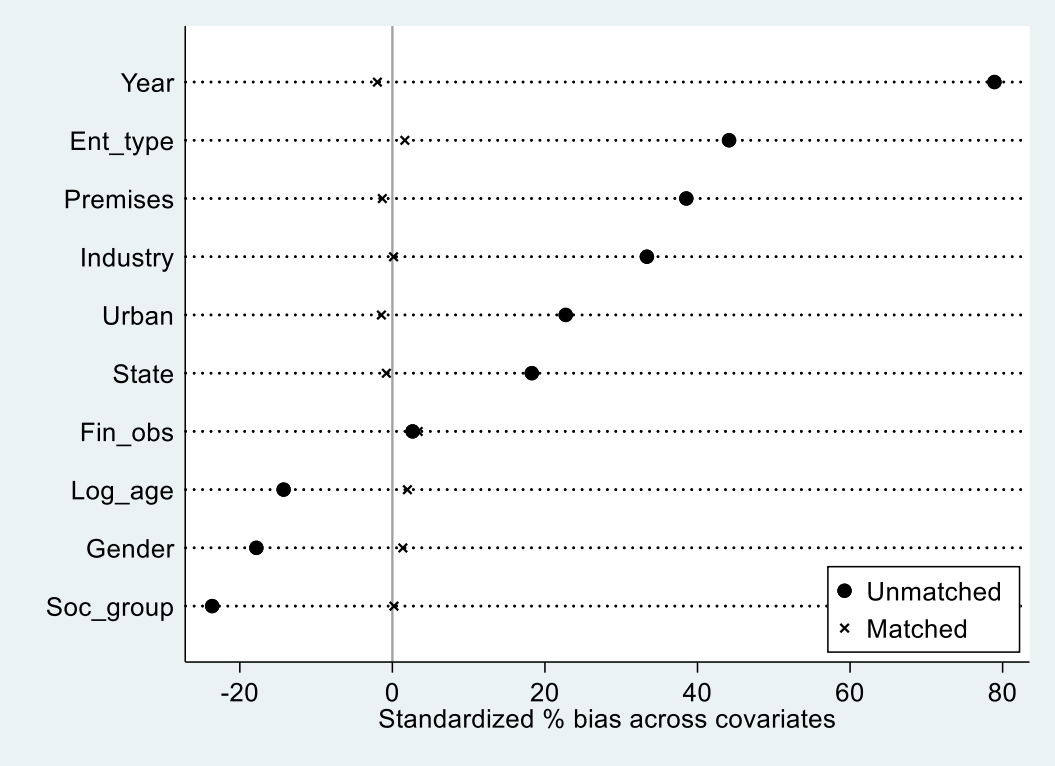
Table 5: Overall matching Balanced test

Sample	Ps R2	LR chi2	p>chi2	Mean Bias	Med Bias	B	R	%Var
Unmatched	0.187	278549	0	29.4	23.2	116.2*	0.45*	100
Matched	0.001	557.85	0	1.4	1.4	6.3	1.1	83

∗ if B>25%, R outside [0.5; 2]

Figure 7: Balanced test graph

[14] Due to software limitations with the *teffects psmatch* implementation (large data), we rely on the widely used psmatch2 routine for nearest-neighbour matching and use *teffects ipw* and *teffects aipw* as robustness checks using inverse probability weighting and augmented inverse probability weighting. Both methods estimate the ATT and yield comparable results, enhancing the robustness of our findings.



Before matching, significant imbalances existed between internet adopters and non-adopters, as evidenced by standardised mean differences (SMDs >10%) and statistically significant t-tests ($p < 0.001$). After matching, covariate balance improves across all observable characteristics thereby confirming that the matching procedure reduces these differences and produces a more comparable sample of adopters and non-adopters.

5.4.1 Subsample analysis: Sector-wise

Next, we carry out a sub-sample analysis based on sectoral classification to understand how the effects of internet adoption vary across sectors. In our sample, manufacturing sector accounts for 30.74 percent of firms, services sector accounts for 41.18 percent, and trading accounts for 26.11[15] percent. India has witnessed rapid growth of the services sector alongside manufacturing (Dehejia & Panagariya, 2014). Available evidence further indicates that formal manufacturing firms record lower productivity than firms in the services sector (Rath, 2018). These differences motivate a sector-specific examination of the association between internet adoption and productivity. Table 6 present the subsample results.

Table 6: Sub-sample analysis by sector

	(1) Manufacturing	(2) Services	(3) Trade
Internet	0.267***	0.287***	0.286***
	(0.016)	(0.011)	(0.014)
Gender	-0.852***	-0.524***	-0.444***
	(0.024)	(0.023)	(0.015)
Internet×Gender	0.049	0.17***	0.244***
	(0.044)	(0.022)	(0.024)

[15] The remaining 1.97% are classified as 'others', which we do not include in the subsample analysis.

OBC	-0.041**	-0.085***	-0.115***
	(0.017)	(0.008)	(0.012)
SC/ST	-0.167***	-0.2***	-0.309***
	(0.021)	(0.012)	(0.019)
Premise: Outside the household	0.334***	0.177***	0.196***
	(0.014)	(0.011)	(0.012)
Sector: Urban	0.243***	0.263***	0.297***
	(0.016)	(0.012)	(0.014)
HWE	0.136***	0.065***	0.133***
	(0.013)	(0.008)	(0.011)
Log age	0.068***	0.032***	0.101***
	(0.011)	(0.005)	(0.006)
Financial obstacle	-0.005	-0.015	-0.047***
	(0.025)	(0.015)	(0.017)
Industry × Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Constant	10.649***	10.889***	10.829***
	(0.032)	(0.02)	(0.021)
Observations	486200	651301	412891
R-squared	0.507	0.347	0.249

Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

The results indicate that in manufacturing sector, the interaction term between internet use and gender is positive but statistically insignificant. In contrast, the interaction coefficient is positive and significant in both services and trading. A plausible reason for the insignificant effect in manufacturing is the relatively low adoption rate among female-owned firms, which limits statistical precision. Chen (2025) observes that the services sector experience stronger digital spillovers than the manufacturing sector in India. Although this observation is drawn for the formal sector, our findings suggests that the same argument can be extended to the informal sector.

5.4.2 Subsample analysis: Usage patterns

The NSSO and ASUSE surveys also provide information on how internet adopting firms use the internet . We perform a subsample analysis of these firms to understand these usage patterns. Earlier regressions capture the impact of internet adoption but do not reveal how digital technologies are used within firms. To address this, we perform a principal component analysis following Higon et al. (2025) and analyse how the resulting components relate to productivity.

We apply PCA to 15 binary indicators of digital activity. These indicators cover online orders placed and received, use of Voice over Internet Protocol (VoIP), getting information from general government organizations, interacting with general government organizations, conducting online financial transactions, providing online customer support, delivering digital products or online services, utilizing web-based tools for recruitment, and digital training. This gives the Digital Usage Index. A second index, the Digital Engagement Index, combines these activities with information on web presence and intranet use. Details of the analysis are provided in the appendix.

Table 10: OLS after PCA for USAGE of internet and computer tools

	(1)	(2)
Dependent variable: Log of Labour productivity	Digital Usage index	Digital engagement index
ICT_index	0.06***	0.071***
	(0.008)	(0.009)
Gender	-0.349***	-0.344***
	(0.029)	(0.029)
ICT_index×Gender	0.105***	0.116***
	(0.014)	(0.016)
OBC	-0.067***	-0.063***
	(0.01)	(0.009)
SC/ST	-0.134***	-0.13***
	(0.014)	(0.014)
Premise: Outside the household	0.269***	0.263***
	(0.031)	(0.031)
Sector: Urban	0.132***	0.131***
	(0.013)	(0.013)
Log age	0.077***	0.077***
	(0.005)	(0.005)
Financial obstacle	-0.019	-0.02
	(0.025)	(0.024)
Constant	11.269***	11.272***
	(0.021)	(0.021)
Year×Industry FE	Yes	Yes
State FE	Yes	Yes
Observations	214220	214197
R-squared	0.268	0.271

Clustered standard errors are in parentheses
*** p<.01, ** p<.05, * p<.1

These usage variables are available only for firms reporting internet or computer access. The resulting sample includes 214220 firms, which shows that only a limited proportion of firms have reported detailed usage information. The results show a significant positive association between digital activity and labour productivity. Higher index values are associated with higher productivity in both specifications. Column 1 uses the digital usage index and Column 2 incorporates the digital engagement index. In both specifications, the coefficients are positive and significant, and the interaction term suggests that female-owned firms benefit positively from ICT use.

5.5 Robustness Check

5.5.1 Alternative measure of productivity: Total Factor Productivity

As a robustness exercise, we use Total Factor Productivity (TFP) as an alternate measure of firm performance. TFP is computed using the gross value-added approach. First, firm-level capital stock (excluding land) is deflated using the Cost Inflation Index (CII). This yields real capital stock for each survey year (2010-11 to 2023-24). TFP is then derived using a growth-accounting or Solow-residual method based on a Cobb-Douglas production framework with variable cost shares, as outlined by Şeker and Saliola (2018). Unlike them, we estimate firm-

level productivity using a value-added production function, where Gross Value Added (GVA) is regressed on capital and labour inputs. The value-added approach is more suitable for informal firms as material input data are often noisy or underreported and vary significantly across industry categories. The methodology is also similar to that of Krishna et al. (2018).

As we use pooled cross-sections, we are unable to implement alternative firm-level TFP measures such as Levinsohn and Petrin (2003). Capital is measured as the total fixed capital stock excluding land, and labour is proxied by total employment including the owner. Total Factor Productivity (TFP) is not used in the main analysis due to difficulties in measuring capital stock for informal firms. In particular, capital stock in informal firms ranges from minimal assets such as basic tools or small machinery in very small firms to substantially larger equipment in upper-tier enterprises. These structural differences can bias TFP estimates We therefore rely on TFP only as a robustness check.

The firm-level observations are summed to create industry-level variables for each year:

$$GVA_{jt} = \sum_{i=1}^{N_{jt}} GVA_{ijt}$$

$$K_{jt} = \sum_{i=1}^{N_{jt}} K_{ijt}$$

$$L_{jt} = \sum_{i=1}^{N_{jt}} L_{ijt}$$

The industry-level TFP is calculated using the Cobb-Douglas production function:

$$GVA_{jt} = A_{jt} K_{jt}^{\alpha} L_{jt}^{\beta}$$

Here, A_{jt} is the Hicks-neutral productivity or the total factor productivity. Taking logs and imposing constant returns to scale (α+β=1) to have a complete technological change gives:

$$\ln\left(\frac{GVA_{jt}}{L_{jt}}\right) = \ln(A_{jt}) + \alpha \ln\left(\frac{K_{jt}}{L_{jt}}\right)$$

The capital elasticity α is estimated using OLS, and the log-TFP for each industry-year is calculated as the Solow residual:

$$\ln TFP_{jt} = lnGVA_{jt} - \hat{\alpha} lnK_{jt} + (1 - \hat{\alpha}) lnL_{jt}$$

The log-TFP is exponentiated to obtain a TFP index in levels for interpretation:

$$TFP_{jt} = \exp(\ln TFP_{jt})$$

We collapse the data to the NIC-year level and compute industry-level averages of establishment characteristics, such as location, gender of ownership, internet adoption, and enterprise age. We then regress industry-level TFP by collapsing then to two-digit level and three-digit level NIC-year level on these indicators as follows:

$$TFP_{jt} = \alpha_1 + \alpha_2 \times \overline{Gender}_{jt} + \alpha_3 \overline{Internet}_{jt} + \alpha_4 \times \overline{Gender}_{jt} \times \overline{Internet}_{jt} + \alpha_5 \times \bar{X}_{jt} + \alpha_6 \times Z_{jt} + \mu_{jt} \quad (4)$$

where $\overline{Gender}_{jt}$ is industry-wise share of female ownership$\overline{Internet}_{jt}$ is industry-wise share of internet adoption, $\bar{X}_{jt}$ is the mean of the independent variables from equation (1), Z_{jt} are year fixed effects, while $\mu_{jt}$is the error term. The coefficient α_3 captures the association

between internet adoption and TFP. The coefficient α_4 captures the interaction term between ownership and internet adoption.

Table 11 reports the TFP regressions estimated at the NIC two-digit and three-digit levels. the estimates of equation 4. The table presents the results for two-digit and three digit NIC level specifications. Internet adoption exhibits a strong positive association with industry-level TFP across both specifications. For NIC two-digit industries,limited variation in female participation (about 11 percent) and the aggregation of heterogeneous firms within broad sectors would not capture the effect of internet adoption rather aggregration at broader industry levels (NIC-three digit, here) would reflect the effect of internet adoption. This pattern is observed in prior research that women tend to be concentrated in a narrow set of industries with low entry barriers (or low capital requirements) such as food processing, apparel, retail trade and other services World Bank, 2021).

Table 11: Regression Analysis using Total Factor Productivity

	(1) TFP (NIC 3)
Internet	0.249***
	(0.057)
Gender	-0.716***
	(0.187)
Internet×Gender	10.305***
	(0.462)
Social group	0.155**
	(0.077)
Urban	0.051
	(0.08)
Premise of enterprises	-0.03
	(0.079)
Type of enterprise	0.151**
	(0.069)
Log_age	0.136**
	(0.06)
Financial obstacle	0.165
	(0.257)
Constant	-0.683***
	(0.227)
Observations	920
R-squared	0.152

Robust standard errors are in parentheses
*** p<.01, ** p<.05, * p<.1

level,

In the NIC-3 digit code level the interaction term is positive and significant. This indicates that the productivity benefits associated with internet use differ by gender within more narrowly defined industries, particularly in those segments where female-owned firms are more active or where digital technologies are more intensively used. The increased significance at finer

levels of industry classification therefore reflects higher within-group variation and the stronger link between digital tools and production processes in specific sub-sectors.

5.5.2 Alternative measure of ICT adoption: Computer adoption

While internet adoption serves as a strong indicator of ICT uptake, computer adoption represents another crucial dimension of ICT integration within firms. To check the robustness of our findings, we use computer adoption as an additional proxy and re-estimate the models using both OLS and IV-2SLS specifications. The results are consistent with our baseline estimations. We present only the IV-2SLS in Table 12.

A potential concern in this analysis is endogeneity, since more productive firms may be more likely to adopt computers, which can lead to reverse causality and biased OLS estimates. To address this, we employ an instrumental variable approach using neighbourhood-level computer adoption as an instrument, following equations 2 and 3. This captures the local diffusion effect of ICT adoption while remaining exogenous to firm-level productivity shocks.

Table 12: Instrument variable analysis for computer adoption

	(1) COMPUTER First Stage	(2) LABOUR PRODUCTIVITY Second Stage	(3) COMPUTER First Stage	(4) LABOUR PRODUCTIVITY Second stage
Instrument: Share of ICT adopters[#] in the same industry-region-year/ Share of ICT adopters[#] in the same industry (other regions)	0.804***		-12.98**	
	(0.0144)		(4.886)	
Computer		0.574***		0.737***
		(0.0612)		(0.167)
Gender	-0.0235***	-0.655***	-0.0245***	-0.651***
	(0.00247)	(0.0167)	(0.00253)	(0.0159)
Computer×Gender	0.653***	0.354***	0.723***	0.235*
	(0.00789)	(0.0484)	(0.00798)	(0.119)
Controls	Yes	Yes	Yes	Yes
Industry Year FE	Yes		Yes	
State FE	Yes		Yes	
Constant	0.00306		0.810**	
	(0.00168)		(0.289)	
Observations	1581540	1581540	1581540	1581540

Clustered standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[#]ICT adopters in this table refer to the computer

The 2SLS estimates confirm a positive and significant association between computer adoption and labour productivity. The interaction terms are also positive indicating that female-owned firms benefit from computer use. Overall, these findings suggest that computer adoption enhances productivity and may also narrow the gender gap in firm performance within the informal sector.

6. Conclusion

This study examines the association between internet adoption, the gender of the owner and productivity differentials in the informal sector. Using a rich pooled dataset of five surveys spanning across 14 years and addressing endogeneity through multiple identification strategies, we report several key results:

First, firms that adopt the internet report higher productivity across all specifications. Secondly, female-owned firms exhibit about 45 to 50 per cent lower productivity than male-owned firms on average, yet this gap narrows substantially when women-led firms adopt digital technologies. The effects are economically meaningful: internet-using female-owned firms achieve productivity levels comparable to the median male-owned firm in our sample. Third, these associations are not driven by simple correlation. Our IV-2SLS estimates, which use variation in regionindustry year adoption as an instrument, along with propensity score matching, support the robustness of the results. The findings remain stable after controlling for firm age, sector, urban location, and size. Fourth, productivity exhibits striking heterogeneity across social groups. Fifth, as an additional exercise, we compute TFP (NIC 3-digit level) and find higher productivity among female owners that adopt the internet.

We also conduct a subsample analysis across broad industry groups. Among manufacturing firms, internet adoption does not appear to significantly enhance productivity, and female-owned firms in manufacturing do not show any significant gains in productivity from internet use. In contrast, internet-adopting firms in services and trade sectors exhibit higher productivity, and this positive effect is particularly pronounced for female-owned firms. These sector differences suggests that women entrepreneurs in service and trade activities may be leveraging internet use more effectively to enhance firm performance. The results remain consistent when computer adoption is used as an alternative measure of ICT adoption.

Contrary to concerns that digitalisation might exacerbate disparities, we demonstrate that basic digital adoption can help mitigate gender gaps in the informal sector. We find that female-owned firms that adopt the internet show marked improvements, which suggests that digital tools can offset structural disadvantages. Productivity also varies across social groups, and older, urban, and externally located firms tend to be more efficient.

The study has a few limitations. First, our data cannot distinguish between types of internet usage (e.g., e-commerce versus information search), leaving the "how" of productivity gains as unknown. Second, the post-2016 period, marked by the internet revolution coupled with demonetisation, may represent a structural break that we cannot fully explore due to the unavailability of panel data.

References

Aboal, D. & Tacsir, E. (2015). Innovation and productivity in services and manufacturing : The role of ICT investment. *MERIT Working Papers*. https://ideas.repec.org/p/unm/unumer/2015012.html

Abramova, N., & Grishchenko, N. (2020). ICTs, Labour Productivity and Employment: Sustainability in Industries in Russia. *Procedia Manufacturing*, *43*, 299–305. https://doi.org/10.1016/j.promfg.2020.02.161

Annual Survey of Unincorporated Sector Enterprises (ASUSE) of 2021-2022 India. (2024). National Sample Survey Office.

https://microdata.gov.in/NADA/index.php/catalog/ENT/?page=1&sort_order=desc&ps=15&repo=ENT

Annual Survey of Unincorporated Sector Enterprises (ASUSE) of 2022-2023 India. (2024). National Sample Survey Office. https://microdata.gov.in/NADA/index.php/catalog/ENT/?page=1&sort_order=desc&ps=15&repo=ENT

Annual Survey of Unincorporated Sector Enterprises (ASUSE) of 2023-2024 India. (2025). National Sample Survey Office. https://microdata.gov.in/NADA/index.php/catalog/ENT/?page=1&sort_order=desc&ps=15&repo=ENT

Bardasi, E., Sabarwal, S., & Terrell, K. (2011). How do female entrepreneurs perform? Evidence from three developing regions. *Small Business Economics*, *37*(4), 417–441. https://doi.org/10.1007/s11187-011-9374-z

Bassanini, A., & Scarpetta, S. (2002). Does human capital matter for growth in OECD countries? A pooled mean-group approach. *Economics Letters*, *74*(3), 399–405. https://doi.org/10.1016/s0165-1765(01)00569-9

Bertschek, I., & Niebel, T. (2016). Mobile and more productive? Firm-level evidence on the productivity effects of mobile internet use. *Telecommunications Policy*, *40*(9), 888–898. https://doi.org/10.1016/j.telpol.2016.05.007

Borowiecki, M., Pareliussen, J., Glocker, D., Kim, E. J., Polder, M., & Rud, I. (2024). *The impact of digitalisation on productivity: Firm-level evidence from the Netherlands*. OECD. https://www.oecd.org/en/publications/the-impact-of-digitalisation-on-productivity-firm-level-evidence-from-the-netherlands_e800ee1d-en.html

Cardona, M., Kretschmer, T., & Strobel, T. (2013). ICT and productivity: conclusions from the empirical literature. *Information Economics and Policy*, *25*(3), 109–125. https://doi.org/10.1016/j.infoecopol.2012.12.002

Clarke, G. R. G., Qiang, C. Z., & Xu, L. C. (2015). The Internet as a general-purpose technology: Firm-level evidence from around the world. *Economics Letters*, *135*, 24–27. https://doi.org/10.1016/j.econlet.2015.07.004

Dutta, N., & Sobel, R. (2016). Does corruption ever help entrepreneurship? *Small Business Economics*, *47*(1), 179–199. https://doi.org/10.1007/s11187-016-9728-7

Fagerberg, J. (1994). Technology and International Differences in Growth Rates. *Journal of Economic Literature*, *32*(3), 1147–1175. https://www.jstor.org/stable/2728605

Gal, P., Nicoletti, G., Renault, T., Sorbe, S., & Timiliotis, C. (2025). *Digitalisation and productivity: In search of the holy grail – Firm-level empirical evidence from EU countries*. OECD. https://www.oecd.org/en/publications/digitalisation-and-productivity-in-search-of-the-holy-grail-firm-level-empirical-evidence-from-eu-countries_5080f4b6-en.html

Gang, I., Natarajan, R. R., Sen, K., & Yun, M.-S. (2022). Does the Gender of the Owner Affect the Productivity of Enterprises in India's Informal Economy? *Journal of Income Distribution*, *31(3-4)*. https://doi.org/10.25071/1874-6322.40548

Grimes, A., Ren, C., & Stevens, P. (2012). The need for speed: impacts of internet connectivity on firm productivity. *Journal of Productivity Analysis*, *37*(2), 187–201. JSTOR. https://doi.org/10.2307/43549652

Hall, B. H., Lotti, F., & Mairesse, J. (2013). Evidence on the impact of R&D and ICT investments on innovation and productivity in Italian firms. *Economics of Innovation and New Technology*, *22*(3), 300–328. https://doi.org/10.1080/10438599.2012.708134

Haller, S., & Siedschlag, I. (2011). Determinants of ICT adoption: evidence from firm-level data. *Applied Economics*, *43*(26), 3775–3788. https://ideas.repec.org/a/taf/applec/v43y2011i26p3775-3788.html

Hitt, L. M., & Brynjolfsson, E. (1996). Productivity, Business Profitability, and Consumer Surplus: Three Different Measures of Information Technology Value. *MIS Quarterly*, *20*(2), 121. https://doi.org/10.2307/249475

Intarakumnerd, P., & Goto, A. (2016). Technology and Innovation Policies for Small and Medium-Sized Enterprises in East Asia. *ADBI Working Paper Series No. 578*. Asian Development Bank Institute. https://doi.org/10.2139/ssrn.2812432

Islam, A. M., & Amin, M. (2023). The gender labor productivity gap across informal firms. *World Development*, *167*, 106229. https://doi.org/10.1016/j.worlddev.2023.106229

Jorgenson, D. W., & Stiroh, K. J. (2000). U.S. Economic Growth at the Industry Level. *American Economic Review*, *90*(2), 161–167. https://doi.org/10.1257/aer.90.2.161

Mallick, S., & Yang, Y. (2013). Productivity Performance of Export Market Entry and Exit: Evidence from Indian Firms. *Review of International Economics*, *21*(4), 809–824. https://doi.org/10.1111/roie.12072

Miller, B., & Atkinson, R. D. (2014). Raising European Productivity Growth Through ICT. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3079844

Mukherjee, S. (2018). Challenges to Indian micro small scale and medium enterprises in the era of globalization. *Journal of Global Entrepreneurship Research*, *8*(1). springer. https://doi.org/10.1186/s40497-018-0115-5

Nucci, F., Puccioni, C., & Ricchi, O. (2023). Digital technologies and productivity: A firm-level investigation. *Economic Modelling*, *128*, 106524–106524. https://doi.org/10.1016/j.econmod.2023.106524

OECD. (2020). OECD Digital Economy Outlook 2020. *OECD Publishing, Paris*. https://doi.org/10.1787/bb167041-en

Oliner, S. D., & Sichel, D. E. (2000). The Resurgence of Growth in the Late 1990s: Is Information Technology the Story? *Journal of Economic Perspectives*, *14*(4), 3–22. https://doi.org/10.1257/jep.14.4.3

Pachouri, A., & Sharma, S. (2016). *ADBI Working Paper Series Barriers to Innovation in Indian Small and Medium-Sized Enterprises* . Asian Development Bank Institute. https://www.adb.org/sites/default/files/publication/189392/adbi-wp588.pdf

Sabarwal, S., & Terrell, K. (2008). Does Gender Matter for Firm Performance? Evidence from Eastern Europe and Central Asia. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1286696

Stansfield, M., & Grant, K. (2003). An Investigation into Issues Influencing the Use of the Internet and Electronic Commerce amoung Small-Medium Sized Enterpirses. *Journal of Electronic Commerce Research*, *4*(1), 15–33. https://discovery.dundee.ac.uk/en/publications/an-investigation-into-issues-influencing-the-use-of-the-internet-

Steinmuller, W. E. (2001). ICTs and the possibilities for leapfrogging by developing countries. *International Labour Review*, *140*(2), 193–210. https://doi.org/10.1111/j.1564-913x.2001.tb00220.x

Torres, J., Maduko, F. O., Gaddis, I., Iacovone, L., & Beegle, K. G. (2021). The Impact of the COVID-19 Pandemic on Women-Led Businesses. *Policy Research Working Paper Series, 9817*. The World Bank. https://ideas.repec.org/p/wbk/wbrwps/9817.html

Unincorporated Non-Agricultural Enterprises (Excluding Construction) - JULY 2010 - JUNE 2011 73 round India. (2011). National Sample Survey Office. https://microdata.gov.in/NADA/index.php/catalog/ENT/?page=1&sort_order=desc&ps=15&repo=ENT

Unincorporated Non-Agricultural Enterprises (Excluding Construction) - JULY 2015 - JUNE 2016 73 round India. (2016). National Sample Survey Office. https://microdata.gov.in/NADA/index.php/catalog/ENT/?page=1&sort_order=desc&ps=15&repo=ENT

Vickery , G., Sakai, K., Lee, I., & Sim, H. (2004). ICT, E-Business and Small and Medium Enterprises. *OECD Digital Economy Papers No. 86*. OECD. https://doi.org/10.1787/232556551425

APPENDIX

Details of Principal Component Analysis

The first PCA identified four principal components with eigenvalues greater than one, which together explain approximately 52.45%[16] of the total variance in firms' online and digital activities, as shown in Table 8. Before rotation, the first component alone accounted for 24.73% of the total variance, indicating a dominant latent dimension of digital activity among firms. However, following a varimax rotation, the variance was more evenly distributed, with the first component as online business and finance explaining 24.73%, component 2 which can be described as e-commerce transactions explaining 10.35%, component 3 can be referred as employment and skills development which explains 9.26%, and lastly component 4 that is government services explaining 8.11% of the total variance, together capturing about 52.45% of the total variation.

To construct a comprehensive measure of firms' digitalization, component scores from the four dimensions were averaged using the mean to create a composite digital index. This index was

[16] Although this does not fulfil the 60% criterion based on the Kaiser criterion, we keep only components with more than 1 eigen value. A scree[lot is added for this analysis Fig A1.

then standardized and subsequently used in regression analysis to examine its association with firm productivity, controlling for gender, social group, location, and other firm characteristics.

Regression-based component scores were generated for each dimension and then combined using a row-mean approach, producing a standardised Digital Usage Index. This composite index was subsequently employed in regression analysis to evaluate its association with firm productivity and other firm characteristics (refer to Table 10, column 1).

Table 8: Variance explained results from the extraction of components for the Digital Usage index

	Initial Eigen values			Rotation sums of squared loadings		
Component	Eigenvalue	Proportion	Cumulative	Variance	Proportion	Cumulative
Comp1	3.28	25.21	25.21	2.60	20.01	20.01
Comp2	1.33	10.26	35.47	1.75	13.44	33.45
Comp3	1.15	8.88	44.35	1.41	10.84	44.29
Comp4	1.05	8.04	52.39	1.05	8.1	52.39
Comp5	0.94	7.25	59.64			
Comp6	0.88	6.75	66.39			
Comp7	0.83	6.37	72.76			
Comp8	0.76	5.88	78.65			
Comp9	0.71	5.45	84.1			
Comp10	0.68	5.24	89.34			
Comp11	0.61	4.7	94.04			
Comp12	0.42	3.24	97.28			
Comp13	0.35	2.72	100			

The analysis for the second index retained four components based on the Kaiser criterion (eigenvalues greater than 1) and confirmed via scree plot inspection. These components collectively accounted for 48.08%[17] of the total variance in the original data. A Varimax rotation was applied to help in the interpretation, resulting in a clear factor structure. Table 9 explains the variance and component extraction.

Table 9: Variance explained results from the extraction of components for the Digital Engagement index

	Initial Eigen values			Rotation sums of squared loadings		
Component	Eigenvalue	Proportion	Cumulative	Variance	Proportion	Cumulative
Comp1	3.58	23.86	23.86	2.44	16.29	16.29
Comp2	1.40	9.3	33.16	1.78	11.84	28.12
Comp3	1.17	7.8	40.96	1.53	10.21	38.34
Comp4	1.08	7.22	48.18	1.42	9.48	47.82
Comp5	0.98	6.56	54.75	1.04	6.93	54.75
Comp6	0.97	6.48	61.23			

[17] We only take components that fulfil the eigenvalue criterion of more than 1. A screeplot is added in appendix Fig A2.

Comp7	0.85	5.7	66.92
Comp8	0.83	5.52	72.45
Comp9	0.80	5.32	77.77
Comp10	0.75	5.02	82.79
Comp11	0.69	4.59	87.39
Comp12	0.62	4.15	91.54
Comp13	0.50	3.33	94.87
Comp14	0.42	2.77	97.64
Comp15	0.35	2.36	100

The six rotated components are interpreted and labelled as follows from Table 9: the first component may be referred to as online business and information access, which has high loadings of interaction and information from websites, emails, government, and general information, as well as customer and financial services. The second component is e-commerce and communication, by activities related to online ordering and selling, delivery services and intranet usage. The third component captured the use of the internet for staff training and development. The fourth component represents e-government and financial services. The fifth component represents internet banking. Subsequently, component scores were predicted for each firm. To construct an overall measure of digitalisation, a composite Digital Engagement Index was created by taking the row mean of the five component scores. This composite index was then standardised to produce a continuous, standardised variable. This variable serves as the primary dependent variable in subsequent regression analyses, providing a multiusage measure of ICT utilisation (refer to Table 10, column 2).

Fig A1: Screeplot for Usage PCA

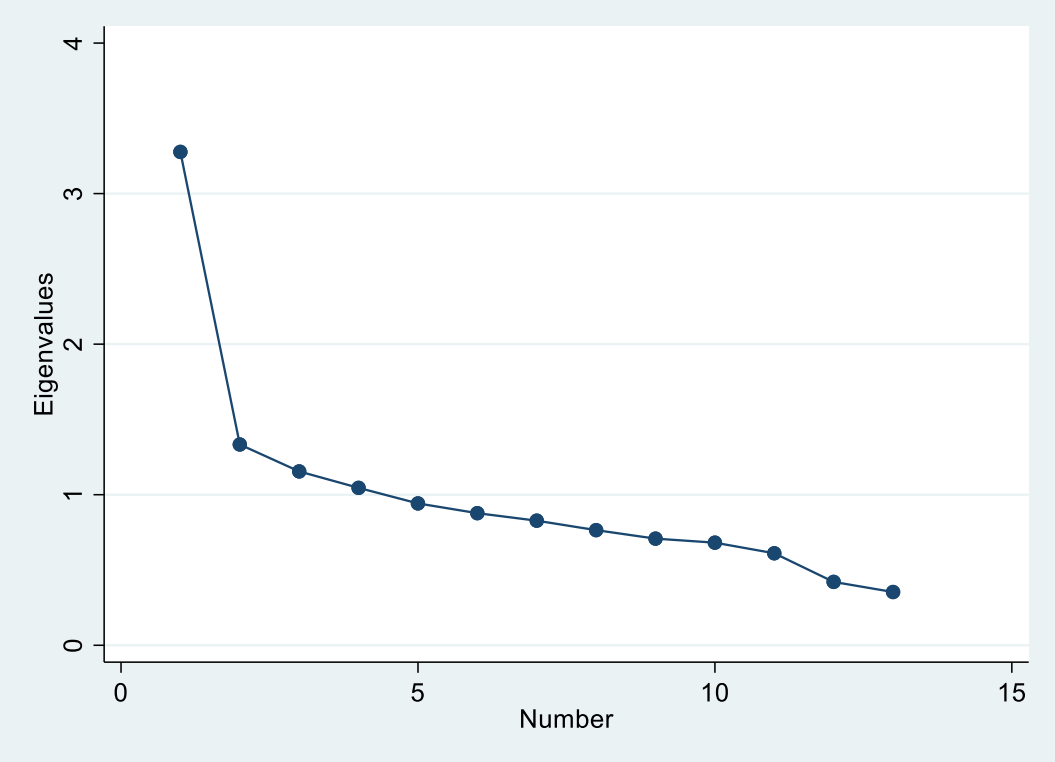
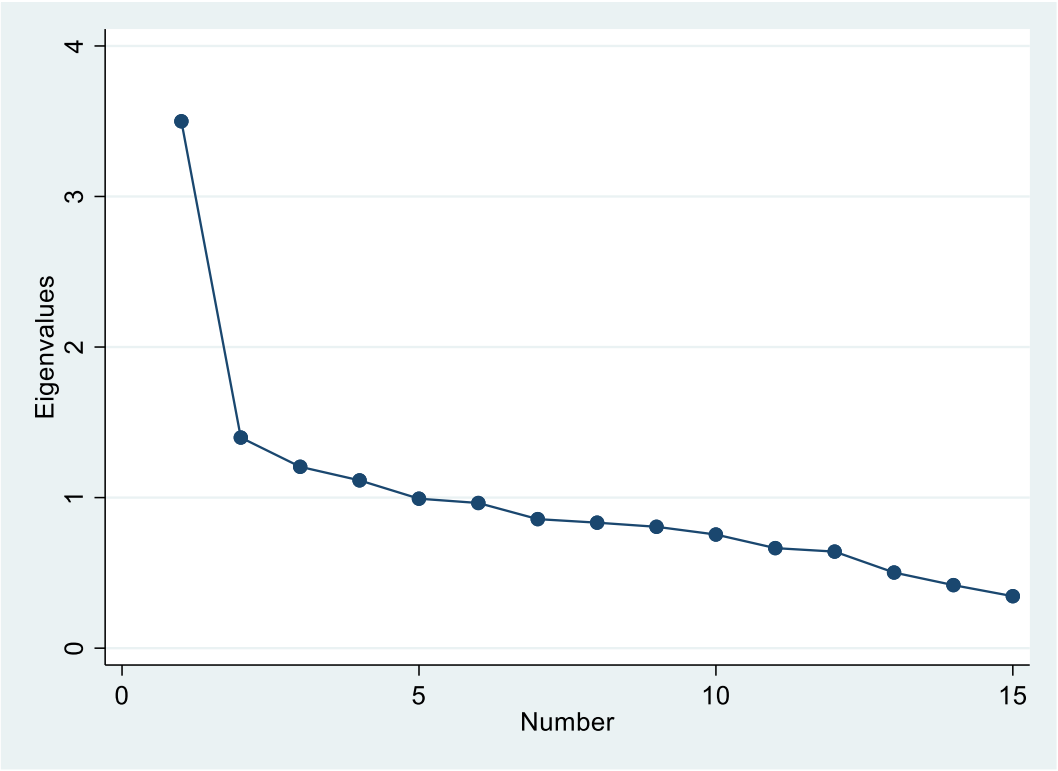


Figure A2: Digital usage index Scree plot


Eigenvalues
0
1
2
3
4
0
5
10
15
Number