Impact of ESG Performance on AI Adoption: Case Study of MSMEs through an Econometric Modelling Approach

Authorship: Vikas Chaudhary[1,2,3], Soumit Pandey[*,4,5,6], Anandajit Goswami[7,8,9]

Author's Contribution

Conceptualization: Anandajit Goswami, Soumit Pandey, Vikas Chaudhary

Methodology: Anandajit Goswami & Soumit Pandey

Software: Soumit Pandey

Validation: Anandajit Goswami

Formal analysis: Vikas Chaudhary, Anandajit Goswami & Soumit Pandey

Investigation: Vikas Chaudhary, Anandajit Goswami & Soumit Pandey

Resources: Vikas Chaudhary, Anandajit Goswami

Data Curation: Vikas Chaudhary, Anandajit Goswami, Soumit Pandey

Writing - Original Draft: Soumit Pandey, Anandajit Goswami, Vikas Chaudhary

Writing - Review & Editing: Anandajit Goswami, Soumit Pandey, Vikas Chaudhary

Visualization: Soumit Pandey

Supervision: Anandajit Goswami

Project administration: Vikas Chaudhary & Anandajit Goswami

Funding acquisition: NA

[1] Affiliation: PhD Research Scholar, Manav Rachna International Institute of Research and Studies
[2] ORCID ID: NA
[3] Email ID: NA
*Corresponding Author
[4] Affiliation: Researcher, Ashoka Centre for a People Centric Energy Transition, Ashoka University
[5] ORCID ID: https://orcid.org/0009-0000-6039-7976
[6] Email ID: soumitpandey.research@gmail.com
[7] Affiliation: Professor, Director, Advisor, MREI, Research Lead, Ashoka Centre for a People Centric Energy Transition
[8] ORCID ID: 0000-0002-2710-0883
[9] Email ID: anandajit@mrei.ac.in; anandajit.goswami@ashoka.edu.in

Declarations

Funding

NA

Conflicts of interest/Competing interests

The authors declare no competing interests.

Data availability

Raw dataset as well as the final/processed dataset will be made available at request.

Code availability

MS Excel and STATA 14 was used for the data analysis of this research. The code file for STATA 14 will be made available at request. The excel sheet will also be made available, both as a source of raw data and final/processed data where the analytics are mentioned.

Ethics statement

This study which involves secondary analysis. Compliance with all academic, ethical, and procedural standards have been ensured.

Consent to participate

NA

Consent to publish

NA

Confidentiality and anonymity

NA

Abstract

The future of VIKSIT Bharat depends on the AI adoption of large and medium, small and micro (MSME) firms because these firms must adapt to the new technological evolution and at the same time generate employment and export potential of India by 2047. Therefore, with this background context, this paper deals with nature, degree and the kind of the impact of Environment (E), Social (S), and Governance (G) performance of 21460 large and MSME firms of Faridabad district of Haryana, India on their artificial intelligence (AI) adoption at a sub-national level.

As no direct observations of the independent (E, S and G) and dependent variable (AI adoption) were available, they were represented by proxy variables. Additionally, independent variables representing company constitution type, rural-urban orientation of the firms, industry type, and geographical area coverage of the firms, were chosen to explain the causality between these independent variables and the nature and degree of AI adoption in these firms.

The research investigates the causal connection using two-stage least squares (2SLS) instrumental variables (IV) estimation model. The model dataset had a challenge of missing data in two key variables, social performance and AI adoption. This was solved using multiple imputation by chained equations (MICE). The missing values were imputed with predictive mean matching (PMM, knn = 5) in a framework of chained equations producing 20 imputed datasets, using the control variables. Before we proceeded with the analysis, the imputations were inspected for convergence between observed and predicted mean of the control variables. We performed the Ordinary least-squares (OLS) and 2SLS estimation with the imputed data after this, which was pooled using Rubin's rules.

In the baseline analysis, OLS was performed with AI adoption as the dependent variable. Based on the result, presence of endogeneity was suspected. To address potential endogeneity within the OLS structure, a 2SLS IV estimation method was performed instrumenting governance performance with industry category (manufacturing/services) and plot size of firm as IVs. The statistical diagnostic tests confirmed that governance of MSME firms measured through institutional variables describing the nature of the firms was endogenous in nature and its casual relation with AI adoption of firms was statistically significant. The tests also showed that the instrumental variables (industry category and plot size), were strong and valid. 2SLS estimates suggested that only social performance and nature/structure/type of the firms were significant in impacting AI adoption. This led to the policy conclusion that to attain VIKSIT Bharat goal, while balancing the technological adoption and evolution of MSMEs, sub-national policies, like, Haryana Enterprises And Employment Policy, 2020, must focus on transparency, regulation, monitoring of the social, governance, and environmental performance of firms in means of digital transformation of MSMEs for the future.

Keywords: Social, Environment, Governance, AI Adoption, Performance, MSME, IV, 2SLS, MICE, Sub-national, Transparency, Regulation

JEL: D9, L1, L2, B4, C5, C1

1. Introduction

Artificial Intelligence (AI) has become an integral part of firm business. Evidence from the market shows a positive association of successful business practices with timely and effective AI adoption strategies (Wamba et al., 2026).

The future of VIKSIT Bharat depends on the adoption of AI of large and medium, small and micro (MSME) firms because these firms must adapt to the new technological evolution and, at the same time generate employment and export potential of India by 2047. In 2025, through various government schemes like PM Vishwakarma Scheme, SFURTI, employment potential of more than 2,20,000 people has been generated. Credit access enhancement for MSMEs is being increased through Credit Guarantee Fund Trust, and the new customised Credit Card scheme for micro enterprises. This is expected to further boost job creation. To achieve Viksit Bharat 2047, it is an imperative for MSMEs to improve their output and job creation potential through skill development, promotion of labour-intensive industries and by making conducive environment for ease-of-business of MSMEs.

Another aspect of firm business strategy that has and is in the process of becoming integral is Environment, Social, and Governance (ESG) parameters (performance) of firms, which is an important criterion to determine their sustainability and ethical practices (Barro et al., 2024; Paužuolienė & Derkach, 2024). A significant link (Xie & Wu, 2025) exists between AI integration and enhancements in ESG performance, especially within the industrial sector, where AI improves, if not optimizes, production efficiency, reduces waste, and aids in adhering to environmental requirements. Predictive analytics can help cut down on carbon emissions, use less energy, and make the supply chains more environmentally friendly for companies and organizations. Huang et al. (2024) also says that AI has improved social and governance aspects by making workplaces safer, getting rid of biases in corporate decision-making, and making corporate supervision stronger through different data-driven methods.

This study, through research conducted on the impact of ESG performance indicators of 21460 large and MSME firms of the Faridabad district of Haryana, on India's AI adoption, arrived at certain directions for future policy conclusions and decision making. The empirical and quantitative research at the subnational level indicates the nature, type, rural/urban orientation, location, and ESG performance of the firms can significantly impact the extent of AI adoption of these firms. This opens out certain subnational realities for a smooth transition of these MSMEs towards the Viksit Bharat Goal of 2047 while maintaining their sustainability pathway paradigm. However, it presented a unique challenge in the database. There were no direct indications/variables of AI adoption and ESG performance. This research tackled these critical deficiencies by formulating a comprehensive methodology for the creation of proxy variables (Frost, 1979; Lubotsky & Wittenberg, 2006) to analyse and quantify the AI adoption and ESG performance of each firm within the dataset.

1.1.Literature Review

The interplay of AI adoption, digital transformation, and ESG performance has emerged as a crucial area of policy inquiry, particularly within the context of MSMEs. Empirical evidence

from recent research indicates a significant positive relationship between AI adoption and corporate ESG performance across multiple dimensions (Zhang and Yang, 2024). Liu et al. (2025) further strengthen this finding. Their research identified that overall ESG performance was substantially improved due to AI adoption, which positively impacted all the three individual pillars, with effects proving most pronounced in large firms. Shen et al. (2025) and Chen and Zhang (2025) demonstrated that such positive affect happens through green innovation, improved internal quantity control, supply chain efficiency, and digital transformation. However, this relationship might not be uniformly positive in every situation. Naveed et al. (2025) discovered that AI adoption was enhanced with strong organizational governance structures, which is mostly viable in large firms only. In this context, MEMEs face substantial barriers to AI adoption and digital technologies. According to Pérez-Campdesuñer et al. (2025), the most critical obstacle to AI adoption among Ecuadorian MSMEs was unavailability of structured databases, followed by limited financial capacity. An analysis by Singh et al. (2022) of Indian MSMEs identified that lack of awareness about benchmarking and lack of management vision were the most critical AI adoption barriers.

Targeted government policies have the power to break down such barriers. In a study by Prasetyo (2020), it was observed that Indonesian government expenditure positively contributed to small and medium enterprises, though effects were insignificant for micro-businesses. A study by Lee and Jung (2025) demonstrated that both, initiatives of buying firm and corporate entrepreneurship, enhanced sustainability innovation management among Korean automotive MSME suppliers, with entrepreneurship accentuating external driver effects. Furthermore, Saini and Kharb (2025) identified that regulatory frameworks, corporate governance reforms, and regulatory efforts acted as strategic determinants for successful ESG adoption in India. Moreover, sectoral variations matter significantly. Rangasamy et al. (2024) identified that ESG integration challenges vary across energy, consumer durables, and pharmaceutical sectors, particularly regarding climate and emissions, and hence requires a more nuanced approach which only AI adoption would bring and improve thereafter. Ravikumar et al. (2025) pointed out a far more challenging scenario. Their study emphasized that the adoption of digital finance, which could potentially help overcome supply-side constraints in post-COVID Karnataka, depends on the digital financial literacy of the population, which remains a significant implementation gap till date in India.

The literature review establishes the existence of a complex and multifaceted relationship between AI adoption, digital transformation, and ESG performance of MSMEs. While substantial evidence suggests positive effects on ESG outcomes from AI adoption, relationships are moderated by organizational capabilities, governance structures, institutional contexts, and sectoral characteristics. Critical barriers like financial constraints, skills shortages, inadequate infrastructure, and lack of awareness, continue impeding adoption, particularly in developing economies. Government support, organizational capabilities, stakeholder pressure, and leadership commitment emerge as crucial enablers. VIKSIT Bharat aims to bridge this gap by fostering a supportive ecosystem through policy, skill development and AI adoption. This would enable operational efficiency (an environmental parameter) using

generative AI, inclusive digital platform (a social development parameter), and data-driven compliance tools (a governance parameter) (NITI Aayog, 2025).

A keyword co-occurrence analysis was performed using VoSviewer based on the abstract text of 62 Scopus-indexed publications, with association strength used as the normalization method, which adjusts for differences in keyword frequencies and captures the relative intensity with which pairs of concepts co-occurs across publications, while total link strength was used to represent the cumulative strength of a keyword's connection within the network. The resulting network reveals two dominant and interrelated thematic clusters. The first cluster centres on MSMEs, emphasizing their AI adoption, reflecting a firm-level and diagnostic orientation of the literature focusing on adoption dynamics and constraints. The second cluster is dominated by ESG performance outcomes, indicating their impact on the MSMEs, which drives AI adoption and digital transformation. The keyword "policymaker" functions as the bridging node between these two clusters suggesting an increasing recognition of the need for external policy and institutional interventions in bringing success to the dynamics of integrating ESG and AI adoption among MSME firms. Figure 1 given below provides the network analysis of these two clusters.

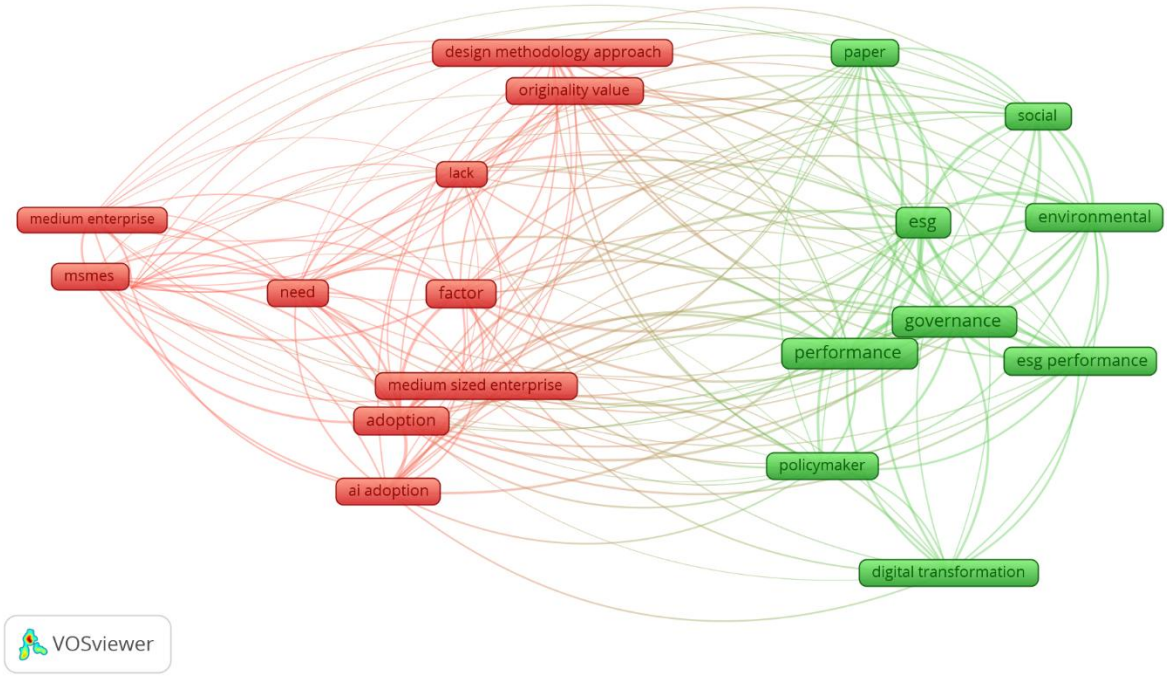


Figure 1: Network Analysis
Source: Author's own work

Based on this literature review this research paper is structured in the following manner. Section 2 highlights the hypothesis, research objectives and research questions of this study. Section 3 highlights the detailed data collection process, sources and methodology adopted for the empirical causal inference based econometric model to test the hypothesis. Further, section 4 deals with the Result and Analysis of empirical models of this study, which further leads to

Discussion in section 5. The paper concludes in section 6, followed by an outlining of the Limitations and Way forward in section 7.

2. Hypothesis, Research Objectives & Research Questions

2.1.*Hypothesis:* AI adoption of MSMEs of Faridabad, Haryana, India increases with ESG performance of the firms.

2.2.*Research Objectives*: (1) To understand the extent and degree of AI adoption of MSMEs of Faridabad with the enhancement of their ESG performance. (2) To understand what future policies can be implemented at a sub-national level to enhance the degree of AI adoption and ESG performance of MSMEs for VIKSIT Bharat 2047.

2.3.*Research Questions:* (1) What is the extent and degree of AI adoption of MSMEs of Faridabad with the enhancement of their ESG performance?
(2) How can future policies facilitate the degree of progressive AI adoption and ESG performance of MSMEs at a sub-national level for VIKSIT Bharat 2047?

3. Data and Methodology

Primary industry level data of 21460 large and MSME firms of Faridabad district, Haryana, India, was used to study the impact of ESG and other firm level parameters on AI adoption of these firms. Due to unavailability of parameters indicating AI adoption and E, S, and G performance of the firms, proxies were considered to reconstruct these key variables. The proxy variables for AI adoption and individual ESG performances were named “aiscore”, "escore", "sscore", "gscore". Further challenges were encountered in the empirical analysis. A significant percentage of missing values were observed in AI adoption and social performance. To tackle this, multiple imputation by chained equations (MICE) were carried out (Wulff & Ejlskov, 2017), where the missing values were imputed with predictive mean matching (PMM, knn = 5) in a framework of chained equations producing 20 imputed datasets, using the observed predictors, including environmental performance, governance performance, and all the control variables. With these imputed datasets, a simple ordinary least squares (OLS) regression (Gaure, 2013) was run with AI adoption as the dependent variable, as a baseline analysis. Presence of endogeneity was suspected from the results of the OLS model. To identify endogeneity in our dataset, the next step was to construct instrumental variables (IVs) (Bascle, 2008; Heng et al., 2019) and running a two stage least squares (2SLS) regression model (James & Singh, 1978; Windmeijer, 2019).

3.1. Calculation of Scores

Multiple variables from the secondary dataset were considered as proxies to construct the variables for individual ESG performance and AI adoption. Table 1 describes the variables considered for environmental performance, social performance, governance performance, and AI adoption. It also describes the other control variables.

Table 1: Description of Variables

Environmental performance	Governance performance	Social performance	AI adoption	Control variables
E1: HSPCB norms (Red, Orange, Green, & White) *Categorical variable*	G1: Registration no. (present or not) *Categorical variable*	S1: Total contractual employees *Quantitative variable*	AI1: Investment in building (in INR lakhs) *Quantitative variable*	Company constitution type (Proprietorship, Partnership, LLP, Pvt Ltd Co., Public Ltd Co., Co-operative, Trust, SHG, LLC (India)) *Categorical variable*
E2: Type of Industry *Categorical variable*	G2: Unit (functioning or closed) *Categorical variable*	S2: Total Casual employees *Quantitative variable*	AI2: Investment in plant & machinery (in INR lakhs) *Quantitative variable*	Area (rural/urban) *Categorical variable*
E3: Type of Product *Categorical variable*	G3: GST number (present or not) *Categorical variable*	S3: Total regular employees *Quantitative variable*		Industry category (manufacturing/services) *Categorical variable*
	G4: Zone (Conforming or non-conforming) *Categorical variable*			Unit category (Large medium, micro, small) *Categorical variable*
	G5: CLU obtained (approved or not approved) *Categorical variable*			Total plot area *Quantitative variable*
	G6: Building plan (approved or not approved) *Categorical variable*			

Source: Author's own work

For the environment score, each variable considered were given a discrete score as per the following:

$$E \in \{0.25, 0.50, 0.75, 1\}$$

This score represents the environmental impact of each firm depending upon the variable. It does not take the value zero since we make a critical assumption that every firm activity leads to some form of environmental degradation.

For the governance score, each variable considered were also given a discrete score, such that:

$$G \in \{0, 1\}$$

A generalized calculation of each score is given below in equation 2.

Equation 1

$$Xscore = \sum_{i}^{n=1} x_i$$

Where, x_i is each variable considered to calculate a score.

3.2. Multiple Imputations by Chained Equations (MICE)

The justification in doing MICE is simply because of a significant proportion of missing values in two of our model's most important variables – AI adoption and social performance. MICE use the following method to mitigate the issue of missingness:

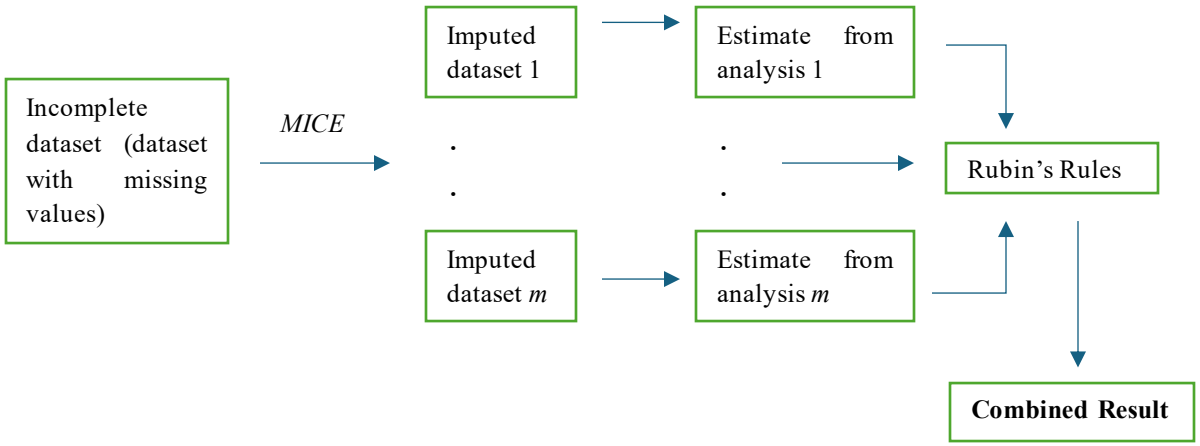


Figure 2: Methodology of MICE

Source: Author's own work

3.3. Ordinary least squares

OLS yields an understanding of a baseline estimate of the relationship between the dependent and the explanatory variables under the assumption of exogeneity (Wooldridge, 2010). The OLS coefficients help us understand the association and causality in the data, which paves the way to address probable endogeneity. Moreover, after completing the MICE process, conducting an OLS was important to check for residual pattern and coefficient stability across all the imputed datasets, as an initial quality check of the imputations, to ensure that the reconstructed dataset is sensible and does not cause implausibility of the parameter values. Equation 3 gives the OLS model run on all the imputed datasets at once.

Equation 2

$$aiscore = \beta_o + \beta_1 escore + \beta_2 gscore + \beta_3 sscore + \beta_4 company\ type + \beta_5 industry\ category + \beta_6 unit\ category + \beta_7 plotsize + \beta_8 area\ category + u_1$$

Since problem of endogeneity was suspected, we proceeded toward a more robust identification strategy, 2SLS since OLS cannot yield unbiased estimates in presence of endogeneity.

3.4. Two-stage least squares/ Instrumental Variables

While STATA allows 2SLS to be run on multiple imputed (MI) data, built-in IV post estimation commands for diagnostic tests like weak instrument, overidentification (validity) and endogeneity, are not available (Mogstad et al., 2021). To overcome this, we computed all the IV diagnostics within of each imputed dataset (m = 20) separately (Enders, 2022; Royston & White, 2011), in a loop command, and recorded them in empty matrices. This method respects the structure of the MI and allows IV diagnostics to be run without error. 2SLS on each imputed dataset, described in equation 4 & 5 below was run multiple times, testing various variables for endogeneity, using multiple IVs. The model ran with governance performance as the endogenous variable, and plot size and industry category as IVs gave the most favourable results.

At first, the m-th imputed dataset is loaded. Then, the first stage of 2SLS is carried out as given below, where we estimate the reduced form equation.

Equation 3

$$gscore_i = \pi_0 + \pi_1 escore + \pi_2 sscore + \pi_3 company\ category + \pi_4 unit\ category + \pi_5 area\ category + \pi_6 plot\ size + \pi_7 industry\ category + v_1$$

From equation 4 we obtain the predicted values $\widehat{gscore}$, which contains only exogenous information. After this, we proceed to the second stage.

Equation 4

$$aiscore = \beta_o + \beta_1 escore + \beta_2 \widehat{gscore} + \beta_3 sscore + \beta_4 company\ type + \beta_5 unit\ category + \beta_6 area\ category + u_1$$

3.5. *Diagnostic Tests*

Diagnostic tests examine the validity and strength of IVs. A good IV should adhere to the below given rules. Given the multiple regression equation is –

Equation 5

$$y_1 = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \cdots + u_1$$

Where, x_1 is the endogenous variable that is correlated with the error term u_1. The other independent variables are exogenous. Now, given that z_1 is the instrument for the endogenous variable x_1, z_1 needs to adhere to the following three fundamental rules –

(1) Instrument $z_1$does not appear in the original regression equation.

(2) $cov(z_1, x_1) \neq 0$

$\Rightarrow x_1 = \alpha_0 + \alpha_1 z_1 + v_1$

(3) $cov(z_1, u_1) = cov(z_1, y_1 - \beta_0 - \beta_1 x_1 - \beta_1 x_2 - \cdots) = 0$

$\Rightarrow cov(z_1, y_1) - cov\beta_1(z_1, x_1) = 0$

$\Rightarrow \beta_1 = \beta_1^{IV} = \frac{cov(z_1, y_1)}{cov(z_1, x_1)}$

Table 2 describes each of the diagnostic tests and their interpretation.

Table 2: Diagnostic tests

Test Category	Specific Test / Statistic	Null Hypothesis H_0	What It Assesses	Decision Rule
Instrument Strength	Minimum Eigenvalue Statistic.	Instruments are not weak.	Detects weak instruments using the Cragg–Donald minimum eigenvalue.	Higher values = stronger instruments.
Overidentification Tests	Sargan χ^2statistic.	Instruments are valid (uncorrelated with error term)	Tests whether instruments satisfy exclusion restrictions.	if $p \leq 0.05$, we reject H_0. It means that at least one instrument is invalid
	Sargan p-value	same as above	same	If $p > 0.05$, it means that instruments are valid.
	Basmann χ^2 statistic	Instruments valid (same as Sargan)	Same as Sargan but robust to certain conditions.	If $p \leq 0.05$, reject H_0.
	Basmann p-value	same	same	If $p > 0.05$, it means that instruments are valid.
Endogeneity Tests (Durbin-Wu-Hausman)	Durbin χ^2 statistic	Regressor is exogenous	Tests whether OLS is consistent.	If $p \leq 0.05$, reject H_0. It means that the variable is endogenous
	Durbin p-value	same	same	If $p > 0.05$. It means that regressor is exogenous.
	Wu–Hausman F statistic	Regressor is exogenous	F-test version of the DWH test	If $p \leq 0.05$, it means the variable is endogenous.
	Wu–Hausman p-value	same	same	If $p > 0.05$. It means the variable is exogenous.

Source: Author's own work

4. Results & Analysis

4.1.Multiple Imputations by chained equations (MICE)

Table 3 given below gives us an understanding of why MICE was necessary. After examining the extent of missingness in the dataset, it is evident that AI adoption had 42.5% and social performance had 10.8% missing values. Given that both AI adoption and social performance are one of the most important variables in this study, such extent of missing data, especially in AI adoption, which is our dependent variable made the complete-case analysis infeasible due to eliminating nearly half the sample and introducing substantial bias. MICE helped reconstruct missing values using predictive relationships among all variables in the dataset. It enabled an efficient and unbiased estimation, while accurately reflecting the uncertainty associated with imputing nearly half of the AI adoption values. It ensured that the OLS as well as the

subsequent 2SLS/IV estimations would be based on a complete dataset, thus ensuring a reliability of statistical tests and empirical results.

Table 3: Results of MICE

Variable	Complete	Incomplete	Imputed	Total
aiscore	12336	9124	9124	21460
sscore	19152	2308	2308	21460

(complete + incomplete = total; imputed is the minimum across m of the number of filled-in observations.)

Source: Author's own work

4.2.Ordinary least squares

The OLS model given in table 4 was estimated on the multiply imputed dataset to provide a baseline understanding of the relationship between the dependent variable, AI adoption, and the independent variables, especially environmental performance, governance performance and social performance. The results indicate that environmental performance has a statistically significant effect on AI adoption, with p-value at 0.011 and a significant positive association with β value at 32.7. Similarly, social performance shows a positive and a significant association with AI adoption, with p-value at 0.001, and β value at 0.021, although the magnitude of the positive association is small. However, interestingly, governance performance shows a statistically significant but negative association with AI adoption. This is quite counterintuitive since it suggests that an improvement in governance will lead to a poor adoption of AI. This motivated us to test for possible endogeneity, especially in governance performance.

As for the other variables related to firm characteristics, most company types have a significant positive effect on AI adoption, indicating that institutional structure influences adoption. Unit category coefficients suggest that medium, small and micro firms have significantly lower AI adoption rates compared to large firms. This is consistent with the analogy that MSME firms often face financial and capacity constraints in their operations. Geographic location appears to be unrelated to AI adoption, suggesting that the adoption rate could be comparable across the spatial classifications.

Table 4: Linear regression

aiscore	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
escore	42.684	9.146	4.67	0	24.735	60.633	***
gscore	-17.458	3.586	-4.87	0	-24.517	-10.398	***
sscore	.021	.01	2.13	.037	.001	.041	**
Company Type	0	.	.	.	.	.	
limited company	290.637	1113.323	0.26	.795	-1938.134	2519.409	
limited liability company	34.889	11.474	3.04	.002	12.399	57.379	***
limited liability partnership	31.612	5.516	5.73	0	20.798	42.426	***
partnership	38.729	4.94	7.84	0	29.04	48.418	***
private limited company	71.451	4.489	15.92	0	62.641	80.261	***
proprietorship	31.837	2.014	15.81	0	27.882	35.792	***
public limited company	94.782	41.583	2.28	.023	13.227	176.337	**

partnership	75.751	22.207	3.41	.001	32.223	119.28	***
self-help group	24.053	2.511	9.58	0	19.129	28.978	***
trust	20.854	3.868	5.39	0	13.271	28.438	***
unit category	0	.	.	.	.	.	
medium	-2782.386	1708.166	-1.63	.104	-6140.879	576.107	
micro	-3578.028	1705.901	-2.10	.037	-6931.744	-224.312	**
small	-3507.336	1705.846	-2.06	.04	-6860.939	-153.733	**
small	-3167.392	1707.06	-1.86	.064	-6523.375	188.59	*
Area category	0	.	.	.	.	.	
urban	.89	4.517	0.20	.844	-7.964	9.744	
Constant	3533.596	1703.963	2.07	.039	183.619	6883.573	**

Mean dependent var	.	SD dependent var	.
R-squared	.	Number of obs	14153
F-test	.	Prob > F	.

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

Source: Author's own work

The result of the diagnostic tests given in table 5 paints a consistent picture. Across the 20 imputed datasets, the results unanimously support the appropriateness of considering the IVs route. The minimum eigenvalue test that examines IV strength consistently falls within the range of 68-69 (average 68.7), substantially surpassing the minimum threshold value of 10 for identification of weak IV, as stated by Staiger and Stock (1997). This confirms that our IVs, industry category and plot size, are strong in the first-stage regression. The Sargan-Basmann overidentification tests (Basmann, 1957; Sargan, 1958) to diagnose the validity of IVs reinforces their appropriateness, as all the of the p-values (average 0.69) and majority of χ^2statistics (average 0.32) exceed the 5% significance level indicating that the null hypothesis of IV validity cannot be rejected. The similarity of Sargan and Basmann p-values across the datasets indicates that the validity of the IVs is robust to potential heteroskedasticity (since Sargan tests validity under homoskedasticity and Basmann under heteroskedasticity). The p-values of Durbin-Wu-Hausman tests (Durbin, 1954; Wu, 1973; Hausman, 1978; Patrick, 2021), that assess whether the suspected endogenous variable (gscore) should be treated as endogenous or not, show mixed results, with majority failing the 5% significance level, while others exceed bringing the average p-value to 0.07. The implications of this are that endogeneity is present, justifying the choice of 2SLS. The F and χ^2 statistics reinforce the justification with all values well above the 5% level of rejecting the null hypothesis, statistically indicating strong evidence of endogeneity. The observed variation in table 5, in the statistics and p-values are expected given the characteristics of MI, which reflects uncertainty arising from the missing data.

Table 5: Diagnostic Test results

Imputation No.	Minimum Eigen	sargan		basmann		Durbin-Wu-Hausman			
		χ^2statistic	p-value	χ^2statistic	p-value	χ^2statistic	p-value	F-statistics	p-value
1	68.464753	0.0307646	0.860767	0.0307212	0.860864	6.837197	0.008928	6.830835	0.008969
2	68.467414	0.0002248	0.988038	0.0002245	0.988046	6.148106	0.013155	6.142086	0.013212
3	68.387204	0.0345133	0.852619	0.0344646	0.852722	4.752028	0.029264	4.746907	0.029367

4	68.392628	0.080997	0.775951	0.080883	0.776104	6.897809	0.00863	6.89142	0.00867
5	68.421036	0.890429	0.345361	0.889227	0.345687	10.61996	0.001119	10.61291	0.001126
6	69.212345	0.000136	0.990667	0.000136	0.990673	1.119771	0.289968	1.118277	0.290308
7	68.416948	0.362665	0.547029	0.362126	0.547308	2.952644	0.085737	2.949087	0.085948
8	68.33975	0.205255	0.650512	0.204968	0.65074	2.423137	0.119555	2.420127	0.119808
9	68.513396	0.195328	0.658518	0.195055	0.658742	5.851166	0.015567	5.845314	0.015631
10	68.471766	0.006435	0.936064	0.006425	0.936109	4.307544	0.037944	4.302766	0.038068
11	68.248751	0.080827	0.77618	0.080713	0.776333	6.178118	0.012934	6.172081	0.01299
12	68.302629	0.003641	0.951996	0.003619	0.95203	4.670303	0.030689	4.665242	0.030796
13	68.528682	0.247133	0.619101	0.246781	0.619346	6.825701	0.008986	6.819344	0.009027
14	68.460375	0.314648	0.574842	0.314211	0.575108	1.817732	0.177584	1.815396	0.177884
15	68.602109	0.247241	0.619024	0.246896	0.619269	5.772206	0.016282	5.766401	0.016348
16	68.412136	2.206778	0.137405	2.204003	0.137653	13.18789	0.000282	13.18154	0.000284
17	68.43002	0.028686	0.865505	0.028646	0.865599	4.819087	0.028146	4.813917	0.028247
18	68.49963	0.695273	0.404376	0.694325	0.404697	0.379513	0.537864	0.378987	0.538156
19	68.443336	0.013246	0.908371	0.013227	0.908435	5.235978	0.022124	5.230514	0.022209
20	68.520565	0.803795	0.369963	0.802692	0.370288	11.49429	0.000698	11.48737	0.000703

Source: Author's own work

Table 6 gives the IV estimation results using 2SLS regression framework over the pooled multiply imputed dataset. In this model, governance performance is considered the endogenous variable, industry category and plot size as the IVs and AI adoption as the dependent variable (as in the case of the OLS model). The results, unlike what we observed in the OLS model, where governance performance had a significant but negative effect on AI adoption, indicate that after correcting for endogeneity, governance performance now has a positive association but a insignificant effect on AI adoption. Thus, there is no statistically reliable evidence that governance causally increases AI adoption in the firms under this study. However, it also implies that the direction of the association was biased due to endogeneity, and the 2SLS estimates are more credible. Similarly, environmental performance, which was significant under OLS, has become insignificant under the IV framework, suggesting that once IVs were used for governance performance and as a result confounding factors controlled for, any apparent association between the environmental performance and AI adoption for the firms under this study that was observed under OLS, is no longer robust to IV correction. However, social performance remains positively associated and statistically significant, implying a consistent and robust relationship between social performance and AI adoption even after addressing endogeneity. Similar to environmental performance and governance performance,

company type becomes insignificant under 2SLS in explaining AI adoption. However, unit category remains highly significant and negatively associated to AI adoption, indicating the persisting scaling disadvantages to MSME firms.

Table 6: Instrumental variables (2SLS) regression

aiscore	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
gscore	72.829	51.718	1.41	.161	-29.361	175.018	
escore	15.943	19.275	0.83	.409	-22.087	53.973	
sscore	.028	.009	3.16	.003	.01	.046	***
Area category	0	.	.	.	.	.	
urban	-.714	22.258	-0.03	.974	-44.339	42.91	
Company Type	0	.	.	.	.	.	
limited company	300.635	852.593	0.35	.728	-1477.997	2079.268	
limited liability company	38.54	128.772	0.30	.765	-213.848	290.928	
limited liability partnership	19.626	120.66	0.16	.871	-216.864	256.116	
partnership	19.115	119.8	0.16	.873	-215.688	253.918	
private limited company	60.571	119.436	0.51	.612	-173.519	294.661	
proprietorship	22.6	119.335	0.19	.85	-211.292	256.493	
public limited company	81.394	120.725	0.67	.5	-155.224	318.013	
partnership	51.195	129.626	0.39	.693	-202.868	305.257	
self-help group	9.299	123.178	0.08	.94	-232.126	250.723	
trust	11.232	137.767	0.08	.935	-258.787	281.251	
industry category	0	.	.	.	.	.	
services	-2777.717	799.468	-3.47	.003	-4450.245	-1105.189	***
unit category	-3556.697	793.08	-4.48	0	-5215.843	-1897.552	***
medium	-3496.375	792.258	-4.41	0	-5153.824	-1838.926	***
micro	-3141.545	802.609	-3.91	.001	-4815.707	-1467.383	***
Constant	3511.907	802.827	4.37	0	1837.457	5186.357	***

Mean dependent var	.	SD dependent var	.	
R-squared	.	Number of obs	14153	
Chi-square	.	Prob > chi2	.	

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

Source: Author's own work

5. Discussion

This study investigated the association between firm level ESG performance and AI-driven technological investment outcomes of firms based in the Faridabad district of the state of Haryana, India, using a cross-sectional dataset characterized by substantial missingness in two of the key variables, AI adoption and social performance of firms. The analysis gives us an understanding of the critical role of data completeness and methodological rigor in accurately estimating relationships within sustainability and technological frameworks. One of the central challenges faced in this research was the lack of standardized and reliable reporting of ESG and AI metrics of the firms (Aziz & Alshdaifat, 2024; Buboi, 2025). As a result, they had to be constructed from the available variables, which lead to environmental and governance parameters having an insignificant effect on AI adoption in the 2SLS model. This limitation directly affects the model stability and restricts the strength of causal inference, emphasizing

the need to improve harmonized ESG reporting requirements and disclosure norms across regulatory agencies and industrial bodies.

To address missing data and reduce potential bias, MICE was applied, which was particularly important because nearly 42.5% of observations for AI adoption and 10.8% for social performance were initially missing. MICE allowed the reconstruction of plausible values based on observed correlations across variables, preserving the statistical power and the generation of robust regression estimates through pooled imputed datasets. Without MICE, the IV estimation approach that followed would not have been feasible due to insufficient sample size.

The results reveal meaningful differences across OLS and 2SLS/IV models. Under OLS social performance demonstrated a significant positive association with AI adoption ($p < 0.05$), indicating firms with stronger employee mix (regular, contractual, and casual) exhibit better technological investment outcomes. We cannot comment on labour conditions of these firms as that was not a variable in aggregating the social performance score. However, we can assume good working conditions in these firms since they are associated with high levels of AI adoption. After correcting for endogeneity through 2SLS, social performance became more strongly significant ($p < 0.01$), and its coefficient also improved from 0.021 to 0.028, reinforcing the robustness of the relationship and suggesting potential causal effect. In contrast, governance and environmental performance, which appeared strongly associated under OLS, became statistically insignificant under 2SLS, indicating previously observed associations may have been driven by confounding factors rather than genuine causal mechanisms. However, given the variables considered in formulating the environment and governance scores, intuitively, they should have had a significant effect on the variables considered to aggregate AI adoption. This leads to the idea that maybe reverse causality is present.

The role of unit category (firm size: large, medium, small, micro) emerged prominently with MSME categories exhibiting consistently negative and highly significant effects relative to large firms in both OLS and 2SLS models, suggesting persistent scale disadvantages among MSMEs in accessing, adopting, and optimizing AI technological investment strategies. Resource constraints, informal nature of labour force, and weak institutional structure and support may be contributing to these barriers. This result aligns with evidence of MSMEs in India, which lag technological upgradation due to systemic bottlenecks rather than inherent inefficiencies (Maheshkar & Soni, 2021). These findings provide a strong empirical basis for policy intervention aimed at strengthening technological innovation and transformation in MSMEs through targeted subsidies, skill development, ecosystem partnership, and easier credit access (Husain & Ali, 2014).

6. Conclusion

This study has attempted to establish a relationship between ESG performance of large and MSME firms and AI adoption. The results determine that after controlling for endogeneity, only social performance and nature/type/structure of firms (that is, whether, large, medium, small, or micro) were able to significantly explain AI adoption of these firms. Given that the

coefficients of MSME firms were negative, only large firms showed progressive AI adoption and indicated causal relationship between ESG performance (especially S) and AI adoption. Needless to say, the MSME firms of India have a long way to achieve VIKSIT Bharat by 2047. It must be remembered that with the onset of AI and other new digital technological evolution, MSMEs must gradually live up to the new realities of next-generation automated technologies. Further, the sustainability indicators of these MSME firms must improve too given that India is taking a leadership role in the global climate action measures through mitigation and adaptation measures. To test this, it needs to be seen how the improvement of ESG indicators can further impact the nature and degree of AI adoption of MSMEs.

The analysis of this study directs us towards the fact that schemes like SFURTI, which promote cluster development, sustainable employment, must gradually reorient themselves towards the enhancement of the social performance of subnational MSMEs. To do that, policies must be implemented to ensure equity funding to these MSMEs so that they can be self-reliant in the social indicator adoption and can further grow with technology adoption with the advent of AI. Today or tomorrow, finally these MSMEs must adopt AI for their own survival, and hence in the initial years, equity funding support for making them self-reliant towards such adoption will help. While promoting equity funding into these MSMEs, what will be important is to understand how the MSMEs can be further strongly integrated into the Make in India campaign. The nature of the MSME sector also must be kept in mind while doing this integration in order to ensure that sustainable employment generation is created from these sectors. Many of these MSMEs are often first-time entrepreneurs and would need sustained support from the Government of India to enhance their ESG performance and AI adoption. In our empirical analysis of the 21460 firms, the strong relationship between the Social Performance of the firm and AI adoption has clearly emerged. Another support structure that needs to be created for the sustained growth and ESG performance of these MSMEs is the procurement of goods and services by central and state government departments from these MSMEs. Such demand-side support will ensure that MSMEs can enhance their ESG performance over the long term, which in essence will also make them more mature in their long-term adoption of AI. Wherever required, in this transition process of the MSMEs, the Prime Minister's Employment Generation Programme (PMEGP) must be tapped for project costs, making it more attractive for setting up new businesses and helping these MSMEs towards their long-term transition. However, currently, MSMEs have been facing the challenge of access to credit, technology, and infrastructure bottlenecks. Moreover, delayed payment addressal must be immediately looked after through effective policy implementation. Once all this is done, India can move towards a successful AI adoption pathway of MSMEs through ESG performance improvement, which in a way in the long run will help in attaining the Viksit Bharat 2047.

Thus, to attain the VIKSIT Bharat goal, while balancing the technological adoption and evolution of MSMEs, sub-national policies must focus on transparency, regulation, and monitoring of the social, governance, and environmental performance of firms in the means of digital transformation of MSMEs for the future.

7. Limitations and Way Forward

Despite important analysis, this study has a few limitations. The secondary dataset lacked ESG and AI adoption variables and had to be constructed using proxies. It also lacked temporal data that prevented dynamic analysis like Difference-in-Difference. Further research would benefit from longitudinal data based on proper variables measuring ESG performance and AI adoption of firms, to understand the causal inferences more appropriately and definitively. Existing literature has emphasized the role of AI, such as data analytics, automation, and decision-support systems, in improving the firms' environmental and sustainability goals, social compliance and efficiency, and governance quality and responsibility, respectively. In contrast, this study explores the reverse relationship, examining how ESG performance, like environmental management practices in terms of commodities, governance structures in terms of responsibility and quality, and compliance orientation in terms of labour practices, shape the adoption and effective integration of the above discussed components of AI among MSMEs. Further research on an evolutionary game-theoretic framework, modelling the strategic interactions between firms and stakeholders (shareholders, regulators, consumers, employees, etc), could theoretically ground our empirical findings and establish the causality this study has tried to prove.

Acknowledgement

We cordially thank all the survey enumerators who have put immense effort on behalf of the state of Haryana to gather the data used in this study. We are grateful to the Directorate of MSMEs, Haryana for their support.

References

Aziz, N.H.A. and Alshdaifat, S.M. (2024), "ESG reporting: Impacts, benefits and challenges", In *Sustainable horizons for business, education, and technology: Interdisciplinary insights* (pp. 69-76). Singapore: Springer Nature Singapore.

Barro, D., Corazza, M. and Filograsso, G. (2025), "Environmental, social, and governance evaluation for European small and medium enterprises: A multicriteria approach", *Corporate Social Responsibility and Environmental Management*, *32*(1), pp.1291-1308.

Basmann, R.L. (1957) "A generalized classical method of linear estimation of coefficients in a structural equation", *Econometrica: Journal of the Econometric Society*, pp.77-83.

Buboi, A. (2025), "Determinants of ESG Reporting Difficulty in the Absence of Global Standardization", *Acta Universitatis Danubius. Œconomica*, *21*(3), pp.59-82.

Chen, R. and Zhang, T. (2025), "Artificial intelligence applications implication for ESG performance: can digital transformation of enterprises promote sustainable development?", *Chinese Management Studies*, *19*(3), pp.676-701.

Durbin, J. (1954), "Errors in variables", *Review of the International Statistical Institute* , pp. 23-32.

Enders, C.K. (2022), *Applied missing data analysis*, Guilford Publications.

Frost, P.A. (1979), "Proxy variables and specification bias", *The review of economics and Statistics*, pp.323-325.

Gaure, S. (2013), "OLS with multiple high dimensional category variables", *Computational Statistics & Data Analysis*, *66*, pp.8-18.

Hausman, J.A. (1978), "Specification tests in econometrics", *Econometrica: Journal of the econometric society*, pp.1251-1271.

Huang, Y., Liu, S., Gan, J., Liu, B. and Wu, Y. (2024), "How does the construction of new generation of national AI innovative development pilot zones drive enterprise ESG development? Empirical evidence from China", *Energy Economics*, *140*, p.108011.

James, L.R. and Singh, B.K. (1978), "An introduction to the logic, assumptions, and basic analytic procedures of two-stage least squares", *Psychological Bulletin*, *85*(5), p.1104.

Lee, S.Y. and Jung, S. (2025), "Internal and External Drivers That Foster Sustainability—Integrated Innovation Management of Micro and SME Suppliers: A Focus on Corporate Entrepreneurship", *Sustainability*, *17*(18), p.8458.

Little, R. and Rubin, D. (1987), "Multiple imputation for nonresponse in surveys", *Wiley*, *10*, p.9780470316696.

Liu, Y., Song, J., Zhou, B. and Liu, J. (2025), "Artificial intelligence applications and corporate ESG performance", *International Review of Economics & Finance*, *104*, p.104559.

Lubotsky, D. and Wittenberg, M. (2006), "Interpretation of regressions with multiple proxies", *The Review of Economics and Statistics*, *88*(3), pp.549-562.

Maheshkar, C. and Soni, N. (2021), "Problems faced by Indian micro, small and medium enterprises (MSMEs)", *SEDME (Small Enterprises Development, Management & Extension Journal)*, *48*(2), pp.142-159.

Mogstad, M., Torgovitsky, A. and Walters, C.R. (2021), "The causal interpretation of two-stage least squares with multiple instrumental variables", *American Economic Review*, *111*(11), pp.3663-3698.

Naveed, K., Farooq, M.B., Zahir-Ul-Hassan, M.K. and Rauf, F. (2025), "AI adoption, ESG disclosure quality and sustainability committee heterogeneity: evidence from Chinese companies", *Meditari Accountancy Research*, *33*(2), pp.708-732.

Aayog, N.I.T.I. (2025), "AI for viksit bharat: The opportunity for accelerated economic growth", *Government of India, Available online: https://www. niti. gov. in/sites/default/files/2025-09/AI-for-Viksit-Bharat-the-opportunity-for-accelerated-economic-growth. pdf (accessed on 2 September 2025)*.

Patrick, R.H. (2021), "Durbin–wu–hausman specification tests", In *Handbook of financial econometrics, mathematics, statistics, and machine learning* (pp. 1075-1108).

Paužuolienė, J. and Derkach, V. (2024), "INTEGRATING ENVIRONMENTAL, SOCIAL AND GOVERNANCE (ESG) PRINCIPLES INTO ORGANISATIONS", *Regional Formation & Development Studies*, *44*(3).

Pérez-Campdesuñer, R., Sánchez-Rodríguez, A., García-Vidal, G., Martínez-Vivar, R. and De Miguel-Guzmán, M. (2025), "Artificial intelligence in Ecuadorian SMEs: Drivers and obstacles to adoption", *Information*, *16*(6), p.443.

Prasetyo, P.E. (2020), "The role of government expenditure and investment for MSME growth: Empirical study in Indonesia", *The Journal of Asian Finance, Economics and Business*, *7*(10), pp.471-480.

Rangasamy, S., Rajamohan, K., Deb, D. and Pillada, N. (2024), "Sectoral analysis of environment social and governance data using deep learning model", *International Journal of Innovation and Technology Management*, *21*(08), p.2440005.

RAVIKUMAR, T., RAGHUNANDAN, G., BENEDICT, J. and KRISHNA, T. (2025), "Digital Financial Inclusion and Financial Resilience of Micro and Small Entrepreneurs in Post-COVID-19 in Karnataka1", *Finance India*, *39*(1).

Royston, P. and White, I.R. (2011), 'Multiple imputation by chained equations (MICE): implementation in Stata", *Journal of statistical software*, *45*, pp.1-20.

Saini, N. and Kharb, R. (2025), "Strategic enablers for ESG adoption: A modified TISM perspective", *Journal of Financial Reporting and Accounting*.

Sargan, J.D. (1958), "The estimation of economic relationships using instrumental variables", *Econometrica: Journal of the econometric society*, pp.393-415.

Shen, L., Li, Z., Liang, Y., Feng, Y. and Zhang, Z. (2025), "Artificial intelligence adoption and corporate ESG performance: evidence from a refined large language model", *Frontiers in Artificial Intelligence*, *8*, p.1691468.

Singh, A., Thakkar, J. and Jenamani, M. (2022), "An integrated Grey-DEMATEL approach for evaluating ICT adoption barriers in manufacturing SMEsAnalysing Indian MSMEs", *Journal of Enterprise Information Management*, *35*(6), pp.1427-1455.

Staiger, D. and Stock, J. (1997), "To econometrica", *Econometrica*, *65*(3), pp.557-586.

Fosso-Wamba, S., Guthrie, C., Queiroz, M.M. and Oyedijo, A. (2026), "Building AI-enabled capabilities for improved environmental and manufacturing performance: evidence from the US and the UK", *International Journal of Production Research*, *64*(2), pp.545-564.

Windmeijer, F. (2019), "Two-stage least squares as minimum distance", *The Econometrics Journal*, *22*(1), pp.1-9.

Wooldridge, J.M. (2010), *Econometric analysis of cross section and panel data*. MIT press.

Wu, D.M. (1973), "Alternative tests of independence between stochastic regressors and disturbances", *Econometrica: journal of the Econometric Society*, pp.733-750.

Xie, H. and Wu, F. (2025), "Artificial intelligence technology and corporate ESG performance: Empirical evidence from Chinese-listed firms", *Sustainability*, *17*(2), p.420.

Zhang, C. and Yang, J. (2024), "Artificial intelligence and corporate ESG performance", *International Review of Economics & Finance*, *96*, p.103713.

Appendix

Codes run on STATA 14

```
clear

// clears any previous analysis - used to start fresh


import delimited "C:\Users\Soumit Pandey\Desktop\personal\AI & ESG\DATASETS\formatted datasheets 3\Faridabad stata .csv"

// imports dataset


rename totalplotareainsqmtrsmandatory plotsize

// renames the variable totalplotareainsqmtrsmandatory to plotsize. Sometimes STATA has problem dealing with vars with long names.


encode companyconstitutiontypeoptional , gen(company_type_num)

encode areamandatory , gen(area_num)

encode industrycategorymandatory , gen(industry_num)

encode unitcategorymandatory , gen(unit_num)

encode plotownershipmandatory , gen(company_constit_num)

// this encodes categorical text variables to numerical variables. Will be helpful later in analysis as STATA will take (k-1) dummies automatically for these categorical variables.


mi set mlong

// since this dataset has large number of missing values, next step is to do multiple imputation (mi). mi predicts missing values using other variables, while keeping uncertainty. It is helpful when missingness is random, which is the case in our dataset (critical assumption).
```

mi register imputed aiscore sscore

// the variables with missing values are registered for mi

mi impute chained (pmm, knn(5)) aiscore sscore = escore gscore totalplotareainsqmtrsmandatory i.area_num i.industry_num i.unit_num , add(20) rseed(12345)

// Here, "mi impute chained" tells STATA to use Multiple Imputation by Chained Equations (MICE). Each variable with missing values is imputed in turn, conditional on the others. The process is repeated until convergence. Next, "(pmm, knn(5)) aiscore sscore" tells that variables aiscore and sscore have missing values. They are continuous, so we impute them using Predictive Mean Matching (PMM). "knn(5)" for each missing value, STATA finds 5 nearest observed cases and randomly draws one to impute. Advantage of this process is that the imputed values are plausible (always actual observed values, not impossible numbers like negative employees). Next, "= escore gscore installedproductioncapacitymetri totalplotareainsqmtrsmandatory i.company_constit_num i.industrycategory i.unitcategory i.plotownership" are the predictors used to fill in missing values. So basically, the idea is that missing aiscore and sscore is estimated from other firm characteristics (ESG, industry, size, etc.). Next, "add(20)" creates 20 complete datasets, each with different imputed values (reflecting uncertainty). Later, when we run regressions, STATA pools the estimates across all 20. Next, "rseed(12345)" sets the random number seed so results are reproducible. Thus, in short: we replace the missing values with plausible values, multiple times, based on the relationships in our data.

asdoc mi estimate: reg aiscore escore gscore sscore i.company_type_num i.unit_num i.area_num, vce(robust), replace

// In this step, we run OLS regression with robust stantard errore. "asdoc" records the the result on a word doc. "replace" replaces the previous word doc with the most recent results.

reg aiscore escore gscore sscore i.company_constit_num i.industrycategory i.unitcategory totalplotareainsqmtrsmandatory

vif

// to check for multicollinearity

matrix mineig = J(20,1,.) // to test stength of IVs, that is, they are correlated to the endogenous variable, a matrix is constructed

matrix sargan = J(20,2,.) // to test for the validity of IVs, that is, they are uncorrelated to the error term, a matrix is constructed # test 1

matrix basmann = J(20,2,.) // to test for the validity of IVs, that is, they are uncorrelated to the error term, a matrix is constructed # test 2

matrix dwh = J(20,4,.) // to test if the endogenous variable taken is in fact endogenous

local i = 1

```
forvalues m = 1/20 {

    use mi_data.dta, clear      // reload MI dataset each time

    mi extract `m'              // now it will NOT give "mi not set"

    quietly ivregress 2sls aiscore (gscore = plotsize i.industry_num) escore sscore i.area_num i.company_type_num
i.unit_num

    // Testing Weak instruments. To test for the strength of the instrumental variables (IV)
    quietly estat firststage
    matrix mineig[`i',1] = r(mineig)

    // Testing validity
    quietly estat overid
    matrix sargan[`i',1]  = r(sargan)
    matrix sargan[`i',2]  = r(p_sargan)
    matrix basmann[`i',1] = r(basmann)
    matrix basmann[`i',2] = r(p_basmann)

    // Testing Endogeneity
    quietly estat endogenous
    matrix dwh[`i',1] = r(durbin)
    matrix dwh[`i',2] = r(p_durbin)
    matrix dwh[`i',3] = r(wu)
    matrix dwh[`i',4] = r(p_wu)

    local i = `i' + 1
}

matrix list mineig
matrix list sargan
```

matrix list basmann

matrix list dwh

asdoc mi estimate, cmdok: ivregress 2sls aiscore (gscore = plotsize i.industry_num) escore sscore i.area_num i.company_type_num i.unit_num, replace

// runs 2SLS