ISSN Online: 3006-4708
SOCIAL SCIENCE REVIEW ARCHIVES ISSN Print: 3006-4694
https://policyjournalofms.com

The Role of AI-Driven Innovation in Green Structural Transformation and Sustainable Economic Development

Ibad Ur Rahman[1], Dr. Samrana Afzal[2], Ubaid Wahab[3], Adam Khan[4]

[1] Department of International Trade, Southeast University, Nanjing, China.
Email: ibadurrahman141@gmail.com
[2] Department of International Relations, Fatima Jinnah Women University, Rawalpindi, Pakistan. Email: samranaafzal@fjwu.edu.pk
[3] Department of Law (LLB), University of Peshawar, Pakistan.
Email: ubaidalamsp@gmail.com
[4] Department of Law (LLB), University of Peshawar, Pakistan.
Email: adamkhanbarikot@gmail.com

DOI: https://doi.org/10.70670/sra.v4i4.2721

Abstract

Artificial intelligence has become central to policy debates on sustainable development, yet the empirical basis for the claim that AI-driven innovation accelerates green structural transformation remains thin and methodologically fragile. This paper makes two contributions. First, it sets out an integrated framework linking AI-driven innovation, green structural transformation and sustainable economic development, and specifies the conditions under which that relationship is expected to operate. Second, it subjects the two-way fixed-effects design routinely applied in this literature to a controlled assessment on a purpose-built panel whose data-generating process is known, so that the behaviour of the estimator can be examined independently of the measurement problems that afflict cross-country patent and development indicators. The panel comprises 40 economic units observed over 15 periods, giving 600 unit-period observations, 560 in one-period-lag specifications and 520 in the two-period-lag specification; models include unit and period fixed effects with standard errors clustered by unit. The one-period-lagged AI coefficient for green transformation is positive but imprecise (0.129; standard error 0.405; p = 0.750). The income models return small negative coefficients (-0.027 and -0.029) that reach only the 10 per cent significance level. Emissions intensity (0.0031; p = 0.530) and employment (-0.0896; p = 0.645) show no statistically distinguishable association. The digital-capacity interaction (-1.184; p = 0.184) is inconclusive, and the two-period lag reverses the sign of the AI coefficient (-0.613; p = 0.224). The contribution is therefore not a verdict on AI. It is a demonstration that sign, precision, lag structure and sample composition must be read together; that a moderation or mediation claim cannot be sustained by imprecise constituent estimates; and that a relationship hypothesised contemporaneously will not necessarily survive estimation with a lagged regressor. The paper closes with a documented protocol for replicating the design on OECD and World Bank indicators.

Keywords: Artificial Intelligence; Green Structural Transformation; Sustainable Economic Development; Fixed Effects; Cluster-Robust Inference; Research Design

Introduction

The pursuit of sustainable economic development requires technological progress to deliver not only higher productivity but also environmental protection and social inclusion. These objectives are connected in ways that are easy to state and difficult to measure. Improvements in production and income may prove impossible to sustain if they intensify environmental

pressure, and they may be socially fragile if large sections of the population are excluded from their benefits. Artificial intelligence has entered this debate because its applications now reach energy systems, industrial processes, agriculture and public administration, and because the speed of its diffusion has outpaced the statistical infrastructure available to evaluate it [1].

The United Nations Conference on Trade and Development argues that AI can accelerate progress toward the Sustainable Development Goals while simultaneously widening inequalities where access to its capabilities remains uneven. Vinuesa in a consensus-based expert assessment, identified published evidence that AI could support 134 Sustainable Development Goal targets while potentially inhibiting 59. Those figures describe potential enabling and inhibiting relationships rather than a measured net effect of adoption, and their coexistence is the analytical problem this literature has to solve: the same technology appears on both sides of the ledger [2,3]. Within this context, green structural transformation offers a perspective that connects technological change to the composition of economic activity rather than to its level. The concept refers to the reorganisation of production, investment, employment and technological capabilities toward lower-carbon, resource-efficient activities. It involves more than improving the environmental performance of individual firms: it concerns industrial upgrading, cleaner energy systems and the emergence of new productive capabilities. Rahman place renewable energy adoption, cleaner production and climate-compatible diversification at the centre of this transformation, and emphasise the role of institutions and coordinated industrial policy in determining whether it occurs [4,5].

Analytical framework and mechanisms

Three mechanisms connect AI-driven innovation to green structural transformation, and each carries a distinct empirical implication. The first is the optimisation of energy systems. AI applications can improve the forecasting of renewable generation, support demand-side management and coordinate systems with variable renewable output. Where these capabilities are deployed, the effect should appear first in the composition of energy supply and then in the emissions intensity of output. Vinuesa identify smart grids and renewable integration as central pathways to affordable energy and climate objectives [7,8]. The second is the reorganisation of production and resource use. UNCTAD documents applications in process optimisation, supply-chain coordination, agriculture and energy management. From a structural perspective, the significance of these applications is not the immediate efficiency gain but the possibility that firms accumulate technological capabilities that allow them to move into new, greener activities. This is the mechanism that connects AI to structural change rather than to marginal productivity, and it is the mechanism least well captured by patent counts, which record invention rather than use [9]. The third is the generation of information for environmental planning. AI can process large and heterogeneous datasets to detect patterns that conventional methods miss, with applications in climate modelling, ecosystem monitoring and satellite-based detection of environmental change. Better information, however, does not guarantee better decisions; its value depends on the capacity of institutions to interpret it, set priorities and act. Taken together, these mechanisms support a distinction that the empirical literature often blurs: adoption is not transformation. A firm may adopt an AI application without materially changing its environmental performance, and an economy may expand its digital sector without producing greener production [10,11,12]. Whether AI-driven innovation contributes to transformation is therefore an empirical question about the composition of activity, not an inference from the rate of adoption. Firm-level evidence that AI investment is associated with growth and product innovation establishes that the technology has economic consequences; it does not establish that those consequences are green.

Conditions, risks and distributional limits

The relationship is conditional rather than automatic, and the conditions are identifiable. UNCTAD treats infrastructure, data and skills as the three leverage points that determine national AI capability. Digital infrastructure determines whether applications can be deployed at all; skills determine whether they can be adapted to local production problems; institutional

quality determines whether the gains are governed and shared. These conditions are not controls to be added mechanically [1,13]. They are the reason why the same technology may produce different outcomes in different settings, and they are the reason why a single average coefficient estimated across heterogeneous economies is likely to describe no economy in particular. Three risks qualify the expected benefits. The first is distributional. If capabilities remain concentrated among a small group of countries and firms, diffusion may reinforce existing inequalities rather than support broad-based development; Vinuesa warn that unequal access to education, computing capacity and technology can deepen disparities within and between countries [14]. The second is environmental. The computation required to develop and operate AI systems consumes electricity, and the emissions consequences depend on the carbon intensity of the supply; where that supply is carbon-intensive, the applications and the infrastructure pull in opposite directions. The third is governance-related. Transparency, accountability, data protection and regulatory oversight determine whether applications serve social and environmental objectives, and effective regulation requires that policymakers understand both the capabilities and the risks of the technology. The framework therefore generates a set of expectations that are directional rather than quantitative, and that require an empirical design capable of separating a technological association from a structural one. That separation is the subject of this paper [15].

Statement of the problem

Existing research establishes two things separately. It shows that AI has substantial potential to support sustainable development, and it shows that green structural transformation depends on cleaner production, industrial upgrading and institutional capacity. What it does not establish is the link: whether AI-driven innovation is associated with a reorientation of productive activity toward greener structures, and under what conditions that association becomes visible in measurable outcomes [16]. Two obstacles stand in the way, and they are different in kind. The first is substantive. AI activity is measured almost exclusively through patent indicators, which record invention rather than use, weight all inventions equally regardless of commercial application, and reflect national patenting conventions as much as national technological effort. An increase in AI patent families may therefore register an increase in inventive activity with no change whatsoever in how production is organized [17]. The second obstacle is inferential. The standard design in this literature regresses an outcome on lagged innovation with unit and period fixed effects. That design removes stable unobserved heterogeneity and common shocks, but it does not remove time-varying confounding or reverse causality, and a one-period lag imposes a temporal structure that the underlying theory rarely specifies. These two obstacles interact in a way that is rarely made explicit. When the measurement is indirect and the timing is assumed rather than derived, a null coefficient is uninformative: it may indicate the absence of a relationship, or it may indicate that the relationship was specified at the wrong lag, measured with the wrong proxy, or estimated with too little within-unit variation to detect [18]. The published literature seldom distinguishes these possibilities, and the consequence is a body of evidence in which statistically insignificant estimates are read as substantive findings in both directions. This paper addresses the problem at the level of design rather than at the level of a single estimated coefficient. It asks what the standard specification can and cannot deliver when the underlying relationship is known, and it specifies the conditions under which an empirical version of the study could legitimately claim to have found something.

Research objectives

The general objective is to examine the role of AI-driven innovation in green structural transformation and sustainable economic development, and to establish what the conventional panel design can support as evidence.

The specific objectives are:

1. To specify the mechanisms through which AI-driven innovation may contribute to cleaner production, resource efficiency and environmentally sustainable industrial upgrading, and to translate each mechanism into a testable implication.
2. To formalise the relationship between green structural transformation and sustainable economic development as separate economic, environmental and inclusion dimensions rather than a single aggregate.
3. To assess whether green transformation can be represented as a pathway linking AI-driven innovation to development outcomes, and to establish what a credible test of that pathway would require.
4. To examine how digital capacity conditions the AI-transformation relationship, and to distinguish a moderation claim from the imprecision of its constituent estimates.
5. To evaluate the sensitivity of conclusions to the temporal specification, the estimation sample and the reporting of model fit.
6. To derive a documented protocol for estimating the same framework on observed cross-country data, including the indicator choices, coverage constraints and inferential caveats that such an extension would face.

Research questions

The study is guided by five questions:

1. Through which mechanisms can AI-driven innovation contribute to green structural transformation, and which of those mechanisms are observable in aggregate data?
2. How should green structural transformation and sustainable development be measured so that the relationship between them is not built into the measurement?
3. Under a two-way fixed-effects specification with clustered standard errors, what does an association between lagged AI innovation and green transformation look like when the underlying relationship is known to exist at a different timing?
4. Does digital capacity condition the AI-transformation relationship, and can that claim be sustained when the interaction and its components are imprecisely estimated?
5. What would an empirical version of this study require in order to support, rather than merely illustrate, a conclusion about AI and sustainable development?

Contribution

The contribution is threefold. Conceptually, the paper integrates AI innovation, green transformation and sustainable development into a single framework while keeping the three development dimensions separate, so that a favourable income estimate cannot be read as evidence of environmental or social improvement. Methodologically, it demonstrates on a panel with a known data-generating process that the standard specification has identifiable failure modes: a contemporaneous relationship estimated with a lag, an interaction interpreted without its constituent precision, a fit statistic inflated by fixed effects, and a change in sample composition mistaken for a change in lag structure. Substantively, it shows why a null result in this literature cannot be read as a finding about the world until those failure modes have been excluded [19]. The paper also has a secondary contribution that is procedural. It specifies, in advance and in tabular form, the indicator set, coverage constraints and inferential conditions that an empirical extension would need. Stating the protocol before the estimates is a discipline; it prevents the variable definitions from being chosen after the results are known, which is the most common route by which weak evidence becomes a strong claim.

Research design and method

Design

The study combines two components that are usually conflated in applied work: a controlled panel analysis in which the data-generating process is known, and a documented protocol for estimating the same framework on observed cross-country data. The two components serve different purposes and are reported separately throughout. The panel is constructed rather than collected. Its function is analytical rather than descriptive: it isolates the behaviour of a two-

way fixed-effects estimator when the structure of the underlying relationship is known, so that the interpretation of sign, magnitude and precision can be examined in the absence of the measurement error, coverage gaps and definitional inconsistencies that affect cross-country patent and development indicators. Conclusions are therefore drawn at the level of specification and inference, not at the level of economies. Section 2.6 sets out what would be required to move from the first component to the second [20]. Two considerations motivate this ordering. First, the empirical literature on AI and sustainability depends on patent-based proxies whose correspondence with AI use is indirect by construction [21]. Second, claims about moderation and mediation in this area are frequently advanced on the basis of constituent regressions that are themselves imprecise. Establishing in advance what a given specification can deliver is therefore not a preliminary to the analysis; it is the analysis, insofar as the object of interest is the evidential value of a design that is already in wide use.

The panel

The panel is balanced, comprising 40 economic units observed over 15 periods, giving 600 unit-period observations. Lagging the explanatory variables by one period reduces the estimation sample to 560 observations; imposing a second lag reduces it to 520. No observations are lost to missing values, so the reductions are mechanical and fully determined by the lag structure. The dimensions were fixed with reference to the properties of the estimator rather than to the size of any real sample. The number of units was set at 40 because cluster-robust inference based on fewer than roughly forty to fifty clusters is known to be unreliable in samples of moderate size [22]. Forty clusters therefore represent the lower bound at which the conventional asymptotic justification can responsibly be invoked, not a comfortable case. This choice is deliberate, and it defines the principal limitation of the exercise: several of the estimates reported in Section 3 would be more precisely estimated in a larger panel, and the paper reads its own imprecision as a property of the design rather than as a property of the world. Units are identified as S01 to S40 and carry no country names, ISO codes, regional attributes or income classifications. This is not cosmetic. Attaching country labels to constructed observations would invite country-level reading of estimates that no country produced, and would allow the discussion to drift from what the specification can support to what a particular economy is like. The identifiers are neutral by design.

The generating process combined a contemporaneous association between AI innovation and green transformation with unit-specific and period-specific components. It imposed no relationship between lagged AI innovation and any outcome. The estimated specifications use lagged regressors. The mismatch between the timing of the imposed association and the timing of the estimated relationship is intentional: it reproduces a practice that is common in applied work, in which a relationship hypothesised contemporaneously is tested with a one-period lag without an explicit justification, and it allows the consequences of that practice to be observed rather than argued [23]. The construction also has features that a real dataset would not. There is no measurement error, no missingness, no attrition, no serial correlation in the errors beyond what the fixed effects absorb, and no cross-sectional dependence. These omissions are the subject of Section 4.9 and are not incidental: they mean that the estimates reported here are, if anything, more precise than the corresponding real-data estimates would be, and that the imprecision documented in Section 3 cannot be attributed to data quality.

Variables

Five constructs are used: an AI innovation index, a green transformation index, a digital capacity index, an investment measure, and three outcome variables (log income, emissions intensity and employment). Each is a constructed index, not an indicator drawn from a statistical database. Table A1 in Appendix A maps each construct to the indicator that an empirical extension would use, so that the gap between the present measures and observed measures is explicit rather than implied. The AI innovation index is strictly positive, right-skewed and scaled to approximate the log-transformed patent intensity used in observed-data work. It is an index of inventive activity, and like any patent-based measure it records invention

rather than adoption, weights all inventions equally, and cannot distinguish general-purpose from environmentally targeted applications [24]. The green transformation index is bounded in the range reported in Table 3.1 and increases with the degree of green structural transformation. It is a composite by construction, standing in for the standardised combination of environmental patent orientation and renewable energy share that the empirical protocol would use. The digital capacity index is a standardised continuous score and is therefore unbounded, so negative values are possible. This is a property of the construction, not a coding error, and it has an interpretive consequence: the index cannot be read as a readiness percentage, and the main effect in an interaction model estimated at a value of zero is a value within the observed range rather than an out-of-sample extrapolation. The investment measure is expressed in percentage-of-GDP-like units, and the three outcomes are expressed respectively in logarithms, in intensity units and in percentage-like units. None of these magnitudes has an external benchmark, and none should be compared with published country statistics.

Specification

The baseline model relates lagged AI innovation to green transformation, controlling for lagged digital capacity and lagged investment, with unit and period fixed effects:

$$Git = \alpha + \beta AIi, t-1 + \gamma 1Di, t-1 + \gamma 2Ii, t-1 + \mu i + \lambda t + \varepsilon it$$

where G denotes green transformation, AI the innovation index, D digital capacity, I investment, μ unit effects, λ period effects and ε the error term. Lagging the regressors establishes temporal ordering in the specification; it does not resolve endogeneity, and Section 4.2 shows that in this design it does not even recover the relationship that was imposed [25].

The development outcomes are modelled separately rather than combined into an index:

$$Yit = \delta + \tau AIi, t-1 + \theta 1Gi, t-1 + \theta 2Di, t-1 + \theta 3Ii, t-1 + \mu i + \lambda t + uit$$

where Y is, in turn, log income, emissions intensity and employment. The interpretation of τ depends on the outcome: a positive coefficient for log income and a negative one for emissions intensity would describe different dimensions of a potentially favourable pattern, and the paper reports them separately rather than aggregating them into a judgement about sustainability.

Moderation is examined by interacting lagged AI innovation with lagged digital capacity:

$$Git = \varphi 0 + \varphi 1AIi, t-1 + \varphi 2Di, t-1 + \varphi 3(AIi, t-1 \times Di, t-1) + \varphi 4Ii, t-1 + \mu i + \lambda t + \eta it$$

The marginal association implied by this specification is

$$\partial Git / \partial AIi, t-1 = \varphi 1 + \varphi 3Di, t-1$$

and it is reported across the observed range of D rather than at selected values. The sensitivity specification replaces the one-period lag with a two-period lag and is otherwise unchanged, so that the difference between the two estimates reflects the lag and the sample together.

Estimation and inference

All models are estimated by the within transformation with unit and period fixed effects. Standard errors are clustered by unit, which is the level at which the treatment is assigned and at which dependence is expected. With 40 clusters, the cluster-robust variance estimator is justifiable but not comfortable: its finite-sample performance deteriorates when the number of clusters is small and when cluster sizes are unequal (Cameron and Miller, 2015). The reported p-values should be read in that light, and the paper is explicit about which conclusions rest on a p-value close to a conventional threshold. Estimation in levels assumes that the within-unit variation in the regressors is stationary over the fifteen periods. If strong trending behaviour were present, a levels specification would be questionable and a first-difference or growth-rate model would be preferred, with its interpretation stated. This diagnostic is specified here and applied in Section 2.7. No instrumental-variable or quasi-experimental strategy is adopted. A credible instrument for AI innovation would require variation in inventive activity that is exogenous to environmental and development outcomes, and no such source is available in this design. Lagged regressors are not treated as instruments; that they precede the outcome is a matter of temporal ordering, not of exogeneity.

Diagnostics

Four diagnostics are applied before the substantive models are interpreted. Within-unit variation in the explanatory variables is examined, because a fixed-effects design that retains little within variation will produce imprecise estimates irrespective of the true relationship. Correlation among regressors is examined for collinearity. The residual structure is examined for serial dependence and heteroskedasticity. The sensitivity of the estimates to influential units is examined, because with 40 clusters a single unit can move a coefficient materially.

Protocol for empirical extension

The second component of the design is a protocol specifying how the framework would be estimated on observed data. It is stated before any estimates are reported, and it is deliberately specific about coverage constraints, because those constraints determine which parts of the framework are testable at all.

Table 2.1. Proposed indicators for an empirical extension

Construct	**indicator**	**Source and constraint**
AI innovative activity	AI-related patent families per million population, inventor-residence attribution with fractional counting	OECD patent statistics; priority-date dating; estimated counts for recent years must be distinguished from finalised counts
Green technological orientation	Environment-related inventions as a share of domestic inventions	OECD environment patent indicators; definition fixed across the sample period
Energy-system change	Renewable energy consumption as a share of total final energy consumption	World Bank WDI; coverage uneven across low-income countries
Productive structure	Manufacturing value added as a share of GDP	World Bank WDI; not a measure of green industrialisation and must not be treated as one
Economic outcome	Real GDP per capita	World Bank WDI; constant-price series with a fixed base year
Environmental outcome	CO2 emissions per unit of GDP, consistently defined	World Bank WDI; series availability differs by country and year
Inclusion outcome	Employment-to-population ratio	World Bank WDI; subject to coverage and definitional change
Digital capacity	Internet users as a share of population; fixed broadband subscriptions	World Bank WDI; both series available for most countries after 2000
Institutional quality	Rule of law and government effectiveness	World Bank Worldwide Governance Indicators; perception-based and available biennially before 2002

Three features of the protocol are worth stating explicitly. First, the AI measure is an invention measure and will be described as such; it does not observe adoption, organisational implementation, or whether a given invention has an environmental application. Second, manufacturing value added is a measure of productive structure and not of greenness; an increase in manufacturing may occur with no improvement in environmental performance, and the two must be analysed jointly rather than conflated. Third, the green transformation index and the development outcomes must not share components, otherwise an apparent relationship can be generated by construction rather than estimated from data.

Robustness plan and what was executed

Seven robustness checks were specified for the framework. Their status in the present study is reported in Table 2.2 rather than deferred to the discussion, because a check that was not performed should be visible as a gap in the design rather than as a footnote.

Table 2.2. Robustness checks: specification and execution status

Check	Status	Consequence for interpretation
Alternative AI measure (share of total patenting rather than population-adjusted intensity)	Not executed	The population normalisation is untested; conclusions are conditional on it
Individual green-transformation components rather than the composite index	Not executed	The index is treated as a single construct; component-specific associations are unknown
Alternative lag lengths	Executed	Reported in Table 3.6; the comparison confounds lag and sample
Common estimation sample across specifications	Not executed	Differences between the one- and two-lag estimates cannot be decomposed
Influence of high-activity units	Not executed	Coefficient stability across units is unverified
Country-group comparisons	Not executed	Heterogeneity across development levels is unexamined
Alternative outcome definitions and additional controls	Not executed	Results are conditional on the reported outcome set

Reverse causality remains plausible in any version of this design: economies with stronger development outcomes or more ambitious environmental strategies may invest more in AI. Unobserved policy changes may influence innovation and environmental outcomes simultaneously. Fixed effects do not remove these problems, and the paper does not claim that they do.

Research ethics and reproducibility

The study uses no human subjects and no personal data. It involves no primary data collection, no interaction with participants and no processing of identifiable information. Were the framework extended to observed data, the sources are public, aggregated and licensed for research use, and the terms of use would be observed in full. Reproducibility is treated as a design requirement rather than a courtesy. The construction of the panel, the estimation of every model and the production of both figures are scripted, so that the descriptive statistics and estimates in Section 3 can be regenerated from the scripts alone. The script set, the variable dictionary and the estimation sequence accompany the manuscript as a reproducibility package.

Limitations of the design

Five limitations follow from the design and are stated here rather than in the conclusion, so that they constrain the reading of Section 3 rather than qualify it afterwards. First, the observations are constructed. The estimates describe the behaviour of a specification, not the behaviour of economies, and no country-specific inference is available from them. Second, the variables are indices rather than indicators. Measurement error, which in patent data is substantial and in emissions data is definitional, is absent from the construction; the estimates are therefore more precise than the corresponding observed-data estimates would be, and the exercise cannot speak to the consequences of measurement error. Third, a single data-generating process is examined. Different assumptions about persistence, heterogeneity and cross-sectional dependence would produce different sampling behaviour, and the conclusions are conditional on the process specified in Section 2.2. Fourth, the panel is balanced and complete. Unbalanced panels, attrition and systematic missingness are the normal condition of cross-country data, and their consequences for this design are not examined. Fifth, the green transformation construct

is broader than the index. Sectoral labour reallocation, supply-chain restructuring and the environmental quality of manufacturing are not represented, and conclusions about structural transformation are therefore conclusions about the index and not about the process.

Results

Analytical sample

The descriptive statistics use all 600 unit-period observations. The one-period-lag models use 560 observations, corresponding to fourteen usable periods for each of the forty units. The two-period-lag model uses 520 observations, corresponding to thirteen usable periods. The emissions and employment models, which include a lagged green transformation term, also use 560 observations. No observations are excluded for missingness, outlier treatment or influential-unit screening, so every reduction in sample size is a mechanical consequence of the lag structure. Four distinct estimation samples are therefore in use, and this matters for comparison. A coefficient estimated on 600 observations, one estimated on 560 and one estimated on 520 are not directly comparable, because they are averaged over different unit-period combinations. The paper reports the sample size alongside every estimate for that reason.

Descriptive statistics

Table 3.1 reports the distributions of the constructed variables. The figures describe the construction, not any economy, and the table therefore reports no external benchmark and permits no comparison with published national statistics.

Table 3.1. Descriptive statistics for the analytical panel (N = 600)

Variable	Observations	Mean	Standard deviation	Minimum	Maximum
AI innovation index	600	1.528	0.487	0.167	3.018
Digital capacity index	600	0.362	0.325	-0.484	1.144
Investment (% of output)	600	22.861	3.187	13.787	31.823
Green transformation index	600	43.413	6.105	26.145	59.545
Log income	600	9.697	0.328	8.794	10.506
Emissions intensity	600	0.492	0.056	0.354	0.636
Employment (% of population)	600	57.336	1.736	51.636	62.476

Note: values are properties of constructed series and carry no external benchmark. They should not be read as national statistics or compared with published indicators.

The AI innovation index averages 1.528 with a standard deviation of 0.487 and ranges from 0.167 to 3.018, so the within-unit variation available to the fixed-effects estimator is moderate relative to the cross-sectional dispersion commonly observed in patent intensity. The green transformation index averages 43.413 with a range of roughly twenty-six to sixty, which is wide enough to support estimation but carries no interpretable scale. The digital capacity index takes negative values because it is a standardised continuous score rather than a bounded readiness measure. Its observed range, from -0.484 to 1.144, means that the interaction model in Section 3.6 evaluates the marginal association of AI innovation at a point inside the data rather than beyond the observed support. The investment and employment variables have percentage-like magnitudes, but they are constructed and their units are nominal. The descriptive statistics establish nothing about relationships. They establish only that the panel contains sufficient within-unit variation for the fixed-effects estimator to be identified, and that the dispersion of the outcome variables is not degenerate.

AI-driven innovation and green transformation

Table 3.2 reports the baseline specification set out in Section 2.4. The lagged AI coefficient is positive, at 0.129, with a clustered standard error of 0.405 and a p-value of 0.750. The 95 per

cent confidence interval extends from approximately -0.66 to 0.92 and therefore includes a wide range of both positive and negative values. The point estimate is in the anticipated direction; the interval is not informative about the direction.

Table 3.2. Lagged AI innovation and green transformation (baseline specification)

Explanatory variable	Coefficient	Clustered standard error	p-value
Lagged AI innovation	0.129	0.405	0.750
Lagged digital capacity	0.201	n.r.	0.792
Lagged investment	-0.089	0.035	0.011
Observations	560		
Unit fixed effects	Yes		
Period fixed effects	Yes		

Note: standard errors clustered by unit. The standard error of the digital-capacity coefficient is not reported separately in the estimation output; its 95 per cent confidence interval is shown in Figure 3.1. The full-model R-squared for this specification is not reported, because it would include the contribution of the fixed effects and would not measure the explanatory value of the regressors.

The digital capacity coefficient is positive, at 0.201, with a p-value of 0.792, and its interval is wide enough to span an order of magnitude in either direction. The investment coefficient is -0.089, with a standard error of 0.035 and a p-value of 0.011; this is the only coefficient in the baseline model that is distinguishable from zero at the 5 per cent level. The interpretation of this result requires care in a specific direction. Investment is negatively associated with the green transformation index conditional on the other regressors. That is a fitted association within a constructed panel in which the transformation index was generated directly rather than produced from investment. It is not evidence that investment slows green transformation, and it is not a substantive finding about capital formation.

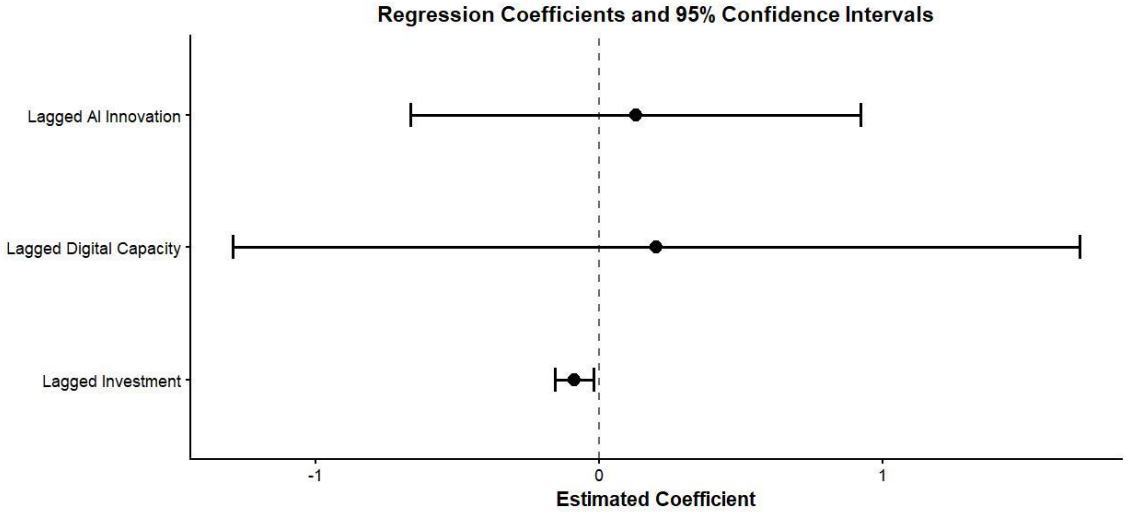


Figure 3.1. Estimated coefficients and 95 per cent confidence intervals, baseline specification. Point estimates and intervals for lagged AI innovation, lagged digital capacity and lagged investment. The dashed line marks zero.

Figure 3.1 presents the same estimates graphically. Two features are immediately legible. The AI and digital-capacity intervals both cross zero with substantial width, which is the visual counterpart of the p-values in Table 3.2. The investment interval is narrow and lies entirely below zero. The figure therefore communicates the central methodological point of this section: the information in a regression table of this kind lies in the width of the interval, not in the sign of the point estimate. A positive coefficient accompanied by an interval spanning a factor of two in each direction supports no directional claim.

AI-driven innovation and income

Table 3.3 reports the income models, estimated with and without the lagged green transformation term. Model 1 includes lagged AI innovation, lagged digital capacity and lagged investment; Model 2 adds lagged green transformation.

Table 3.3. Lagged AI innovation and log income, with and without lagged green transformation

Explanatory variable	**Model 1**	**Model 2**
Lagged AI innovation	-0.0272 †	-0.0290 †
Lagged digital capacity	0.0229	0.0221
Lagged green transformation	—	0.0005
Lagged investment	-0.0027 †	-0.0027 †
Observations	560	560
Unit fixed effects	Yes	Yes
Period fixed effects	Yes	Yes

Note: standard errors clustered by unit. † denotes statistical significance at the 10 per cent level as reported in the estimation output; coefficients without a marker are not significant at that level. Approximate 95 per cent intervals appear in Figure 3.2. Exact standard errors for these two specifications are not reported in the source output and should be regenerated from the accompanying script before submission.

The AI coefficient is negative in both models, at -0.0272 and -0.0290, and declines by 0.0018 when the transformation term is added. Both estimates reach the 10 per cent level and neither reaches the 5 per cent level. The digital capacity coefficient moves from 0.0229 to 0.0221 and is not distinguishable from zero in either specification. Investment is stable at -0.0027. Lagged green transformation enters at 0.0005 and is statistically indistinguishable from zero. The stability of the AI coefficient across the two models is the substantive point. If green transformation were the channel through which AI innovation reaches income, adding the mediator to the specification should attenuate the direct coefficient. A change of 0.0018 on a base of 0.0272 is not an attenuation; it is sampling noise. The finding therefore provides no support for the proposed pathway through green transformation, and Section 3.8 explains why this is not the same as a test of mediation. The marginal significance of the AI coefficient requires a caution that the source output alone does not supply. With forty clusters, the cluster-robust variance estimator is imprecise, and a p-value between 0.05 and 0.10 in a sample of this structure is not a reliable indicator of a non-zero population association. The paper reports the estimate because it was computed, not because it supports a conclusion.

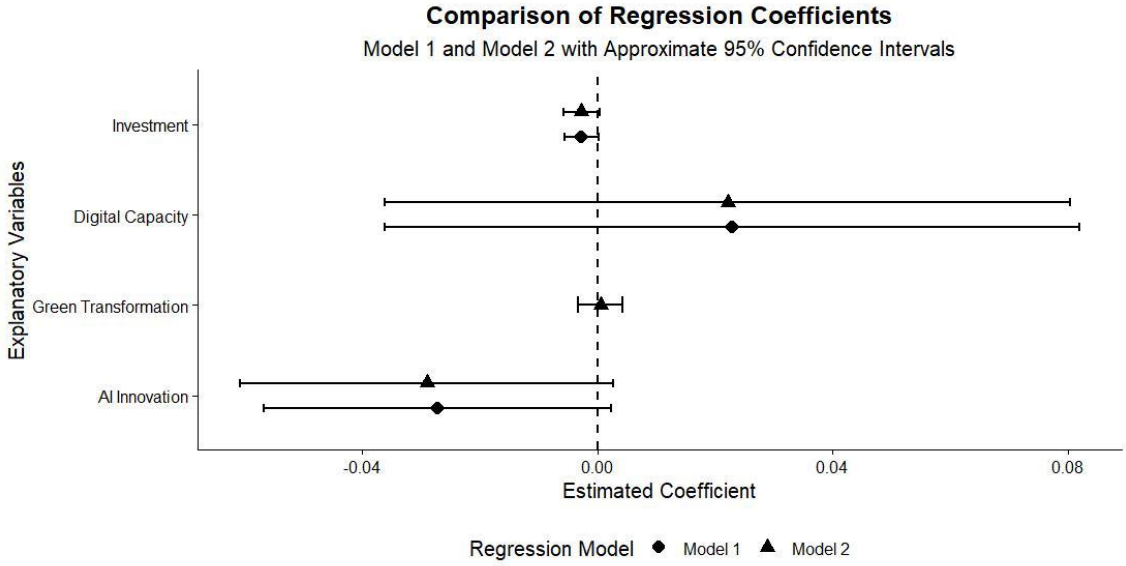


Figure 3.2. Estimated coefficients for the log-income models, with and without lagged green transformation. Model 1 excludes the transformation term; Model 2 includes it. Intervals are approximate 95 per cent confidence intervals.

Figure 3.2 shows why the stability of the AI coefficient should be read as an absence of attenuation rather than as robustness. The two AI intervals overlap almost entirely, which is consistent both with a genuine null effect and with an effect too small to detect at this sample size. The intervals for digital capacity are wide enough to span the entire plausible range of the coefficient. The transformation coefficient is small and precisely centred near zero, which is the visual signature of a variable that carries no information about the outcome in this specification.

Emissions intensity and employment

Table 3.4 reports the two development outcomes that correspond to the environmental and inclusion dimensions. Each model includes lagged AI innovation, lagged green transformation, lagged digital capacity and lagged investment, with unit and period fixed effects.

Table 3.4. Lagged AI innovation, emissions intensity and employment

Explanatory variable	**Emissions: coef. (SE)**	**p**	**Employment: coef. (SE)**	**p**
Lagged AI innovation	0.0031 (0.0050)	0.530	-0.0896 (0.1949)	0.645
Lagged green transformation	0.0003 (0.0005)	0.632	-0.0253 (0.0217)	0.244
Lagged digital capacity	-0.0044 (0.0068)	0.521	0.4046 (0.2763)	0.143
Lagged investment	0.00005 (0.00044)	0.915	0.0252 (0.0162)	0.121
Observations	560		560	
Unit and period fixed effects	Yes		Yes	
Full-model R-squared	0.720		0.615	

Note: standard errors clustered by unit, in parentheses. The full-model R-squared includes the contribution of the unit and period fixed effects and does not measure the explanatory value of the regressors.

In the emissions model no explanatory variable is distinguishable from zero at the 5 per cent level. The AI coefficient is 0.0031 with a standard error of 0.0050 and a p-value of 0.530. Green transformation enters at 0.0003, digital capacity at -0.0044 and investment at 0.00005, all with p-values above 0.5. The appropriate conclusion is narrow: the fitted model does not identify an emissions-reducing association. It does not establish that none exists, and it does not establish that AI or green transformation lacks environmental value. In the employment model the AI coefficient is -0.0896 with a standard error of 0.1949 and a p-value of 0.645. Digital capacity enters at 0.4046 and investment at 0.0252, neither reaching the 5 per cent threshold; green transformation enters negatively at -0.0253 and is equally inconclusive. The model therefore does not distinguish an employment gain from an employment loss, and the point estimate is small relative to its own standard error by a factor of more than two. The two outcomes are reported side by side rather than combined, and that arrangement carries a substantive point. An income result, an environmental result and an employment result answer different questions. A favourable estimate on one dimension would not establish that development had become environmentally sustainable or socially inclusive, and an unfavourable estimate on one dimension would not establish that the others were adverse. The three dimensions of the framework are therefore kept separate in the reporting as well as in the specification. The full-model R-squared values of 0.720 and 0.615 require the same caution as any fit statistic in a fixed-effects model. They include the contribution of forty unit effects and fifteen period effects, which together absorb most of the variation in any panel of this structure. They therefore describe the fit of the full model, including its fixed effects, and not the explanatory contribution of the regressors. A high R-squared in this setting is evidence that the fixed effects are doing their work, which is not in question.

Digital capacity as a moderating condition

Table 3.5. Interaction between AI innovation and digital capacity

Explanatory variable	Coefficient	Clustered standard error	p-value
Lagged AI innovation	0.536	0.480	0.264
Lagged digital capacity	1.941	1.624	0.232
AI innovation × digital capacity	-1.184	0.891	0.184
Lagged investment	-0.094	0.035	0.007
Observations	560		
Full-model R-squared	0.800		

Note: standard errors clustered by unit. Unit and period fixed effects are included. The R-squared includes the contribution of the fixed effects.

The interaction coefficient is -1.184 with a clustered standard error of 0.891 and a p-value of 0.184. Its 95 per cent confidence interval extends from approximately -2.93 to 0.56. The estimate is therefore not distinguishable from zero at the 5 per cent level, and its magnitude is large relative to the main effect, which is the configuration in which interaction models most easily generate spurious conclusions.

The implied marginal association is

$\partial G_{it}/\partial AI_{i,t-1} = 0.536 - 1.184 D_{i,t-1}$

which declines as digital capacity rises and becomes negative at values of D above approximately 0.45. Two conclusions must be separated here. The first is that the estimated interaction has the opposite sign to the one the framework anticipates: if digital capacity were a complement to AI innovation, the coefficient would be positive. The second is that the estimate is too imprecise to support either a positive or a negative moderation claim, and the negative sign in a sample of 560 observations with 40 clusters is not a finding about complementarity. The main AI coefficient of 0.536 is the association at a digital capacity value of zero, which lies inside the observed range but is not the sample average. It is not an unconditional average effect of AI innovation, and it should not be reported as one. This is a general property of interaction models, and it is the reason why marginal effects are reported across the distribution of the moderator rather than at a single point.

Sensitivity to the lag specification

Table 3.6 compares the baseline green transformation model with an otherwise identical specification using a two-period lag of AI innovation.

Table 3.6. Sensitivity of the AI coefficient to the lag specification

Specification	AI coefficient	Clustered standard error	p-value	Observations
One-period-lagged AI	0.129	0.405	0.750	560
Two-period-lagged AI	-0.613	0.505	0.224	520

Note: both models include lagged digital capacity, lagged investment, unit fixed effects and period fixed effects; standard errors clustered by unit.

The AI coefficient changes sign, from 0.129 to -0.613, when the lag is extended by one period. Neither estimate is statistically significant at the 5 per cent level. The 95 per cent confidence interval for the two-period estimate extends from approximately -1.60 to 0.38.

The sign reversal is the most consequential result in this section, and it has two causes rather than one. The first is the lag structure: the relationship imposed in the construction is contemporaneous, so neither the one-period nor the two-period lag recovers it. The second is sample composition: introducing the additional lag removes 40 observations, one period for each unit, and the estimates are therefore not computed on the same unit-period combinations. Comparing specifications that differ in both lag and sample cannot attribute the difference to either. A common-sample comparison, retained in Table 2.2 as an unexecuted check, would be

required to separate them. The practical implication is that a sign change of this magnitude, reported without a common-sample restriction, is not evidence of instability in the relationship. It is evidence that the specification is not identified in a way that supports a directional conclusion at all.

Assessment of the proposed pathway

The framework proposed that green structural transformation might provide a pathway connecting AI innovation to development outcomes. The reported estimates do not establish that pathway, and the formal procedure that would be required to test it was not implemented. Three pieces of evidence bear on the question, and none of them constitutes a test. The association between lagged AI innovation and green transformation is imprecise (Table 3.2). The association between lagged green transformation and log income is negligible and imprecise (Table 3.3). Adding the transformation term to the income model leaves the AI coefficient essentially unchanged, which is the pattern that would be observed if there were no pathway and also the pattern that would be observed if the pathway were too small to detect at this sample size. No indirect effect was computed, no bootstrap confidence interval was constructed, and no product-of-coefficients estimate is reported. This restraint is not a formality. A mediation claim requires an estimate of the indirect effect with a confidence interval, and in the observational case it also requires assumptions about confounding of the exposure-mediator, mediator-outcome and exposure-outcome relationships that fixed effects and temporal ordering do not deliver (Imai, Keele and Tingley, 2010). Presenting three imprecise constituent regressions as a pathway analysis would convert an absence of evidence into an apparent finding, and it is precisely the practice this paper is intended to make visible.

Estimates not reported

Several analyses specified in the framework were not performed, and their absence defines the boundary of this section. The alternative AI measure based on the share of total patenting, the component-specific green transformation models, the common-sample comparison across lag structures, the influential-unit analysis, the development-level subsamples and the alternative outcome definitions all require variables that the constructed panel does not contain. Institutional quality and human capital were specified as moderators and were not tested; only digital capacity was examined. The consequence is that the results in this section are conditional on a single specification family, a single index construction and a single data-generating process. They are reported at that level of generality and no other.

Discussion

Principal findings

Five results emerge, and they are results about a design rather than about economies. First, the baseline association between one-period-lagged AI innovation and green transformation is positive and imprecise: 0.129 with a standard error of 0.405. The point estimate lies in the anticipated direction, and the interval is wide enough to accommodate both a substantial positive relationship and a substantial negative one. Neither reading is supported. Second, the income models produce small negative coefficients of -0.027 and -0.029 that reach the 10 per cent level and not the 5 per cent level. The stability of these coefficients across specifications that include and exclude the transformation term is not evidence of robustness, because the change is too small to constitute attenuation. Third, the emissions and employment models identify no statistically distinguishable association for any explanatory variable, and the two outcomes behave in ways that preclude a joint conclusion about sustainable development. Fourth, the interaction between AI innovation and digital capacity is negative and imprecise, and its magnitude relative to the main effect is large enough that neither sign should be read as a finding about complementarity. Fifth, the AI coefficient changes sign between the one-period and two-period lag specifications, and the comparison confounds the lag with the sample. The reversal is therefore not evidence of instability in the relationship but evidence that the specification does not identify a direction.

Why the null result is not a finding about AI

The most important interpretive point concerns the relationship between what was imposed in the construction and what was estimated. The construction imposed a contemporaneous association between AI innovation and green transformation. Every estimated model used a lagged regressor. No model recovered the imposed relationship, and none could have: a one-period lag estimates the association between AI innovation in period t-1 and transformation in period t, which is a different relationship from the one embedded in the data. This is not an artefact of the simulation. It is the standard condition of the empirical literature, in which a relationship that theory describes as operating through current investment, current deployment or current organisational change is tested with a one-period lag chosen for convenience and justified by the phrase 'to mitigate simultaneity'. The consequence is that a null coefficient in such a specification is consistent with three quite different states of the world: no relationship; a relationship at a timing the specification does not capture; and a relationship too small to detect given the available within-unit variation. The paper can distinguish these states in its own panel because the construction is known, and it cannot distinguish them in any observed panel. The transferable lesson is therefore procedural rather than substantive. A null coefficient should not be reported as evidence of absence unless the lag structure has been justified against the mechanism it is meant to represent, and unless the alternative timings have been estimated on a common sample. The second interpretive point concerns measurement. The AI index records inventive activity, not deployment. A country may increase its AI patent output without changing how its industry uses energy, and a firm may deploy AI extensively without patenting anything. The framework distinguishes adoption from transformation for this reason, and the empirical protocol in Table 2.1 preserves the distinction by describing the patent measure as a measure of invention. An estimate relating invention to transformation is not an estimate of the effect of AI on transformation, and the gap between the two is the central unsolved measurement problem of this literature.

The income result and the limits of marginal significance

The income models return negative coefficients that reach the 10 per cent threshold. There are two temptations here and both should be resisted. The first is to read the negative sign as evidence that AI innovation reduces income. It does not. The coefficient is small relative to its interval, and the 10 per cent threshold in a design with forty clusters is not a reliable boundary. Cameron and Miller (2015) and MacKinnon and Webb (2017) show that cluster-robust inference with few clusters rejects too often, so a marginal p-value in this setting is more likely to be a false positive than in a large-sample design. The second temptation is to read the result as evidence of a trade-off between income and environmental outcomes. It does not support that either, because the emissions model identifies no association at all and the employment model identifies none. A trade-off requires two estimable relationships with opposite signs, not one marginal estimate and two nulls. What the income result does support is a statement about reporting. Coefficients that clear a conventional threshold in one specification and fail it in another, in a design whose inferential properties are marginal, should be reported with their intervals and with the number of clusters, and should not be summarised in an abstract as a directional finding. The paper reports them because they were computed, and the reporting is deliberately flat.

The three development dimensions

The framework treats economic, environmental and inclusion outcomes as separate dimensions, and the results justify that choice. Had the three outcomes been aggregated into a composite sustainability index, the negative income coefficient, the null emissions coefficient and the null employment coefficient would have been combined into a single ambiguous number whose sign would depend on the weights chosen, and whose interpretation would depend on assumptions that no one had tested. The separation also clarifies what a positive result on one dimension would have meant. A positive income association would not have established that development had become environmentally sustainable; a negative emissions

association would not have established that the gains were inclusive. The literature on AI and the Sustainable Development Goals is explicit that the technology operates on both sides of the ledger simultaneously (Vinuesa et al., 2020), and a composite index is precisely the instrument that conceals this. The environmental dimension carries an additional distinction that the aggregate design cannot capture. The environmental benefit of applying AI and the environmental cost of operating AI systems are different quantities, and they move through different channels. The present panel contains no measure of computing-related electricity demand, of the carbon intensity of electricity supply, or of the difference between environmentally targeted and general-purpose applications. It therefore cannot speak to whether the resource requirements of digital infrastructure offset the benefits of its applications, which is the question Vinuesa et al. (2020) identify as central.

Moderation, interaction and the interpretation of imprecision

The interaction result is the clearest illustration of a reporting problem that recurs across this literature. The estimated interaction is negative, large in magnitude and statistically inconclusive. Two conclusions have to be held apart: the estimate does not support the anticipated complementarity between AI innovation and digital capacity, and the estimate does not support the opposite conclusion either. The confidence interval spans values that would imply strong complementarity and values that would imply strong substitution. The main effect of 0.536 is the association at a digital capacity of zero, which is inside the observed range because the constructed index takes negative values, but is not the average. Reporting it as an average effect would misstate the model. This is not a peculiarity of the present index: the same issue arises with any standardised moderator, and it arises in observed data whenever a readiness index is centred at a value that is not the sample mean. The wider point is that moderation claims are frequently advanced on the basis of an interaction coefficient whose sign is read and whose interval is not. Where the interaction is estimated on the same limited within-unit variation as its components, and where the number of clusters is at the lower bound of acceptable inference, the interval is the informative object. The framework's expectation that digital capacity strengthens the AI-transformation relationship remains a hypothesis, and the present design neither confirms it nor refutes it.

The pathway question

The proposed pathway from AI innovation through green transformation to development outcomes remains unresolved, and the paper is explicit that no test of it was performed. The constituent estimates are uninformative individually, and their product would be uninformative jointly: an indirect effect estimated as the product of two imprecise coefficients has an interval wider than either component, and reporting it as a pathway estimate would misrepresent its content. Two conditions would be required for a credible test, and neither is satisfied here. The first is temporal ordering that matches the mechanism, which in this design would mean imposing the association at the lag that the specification estimates. The second is an identification strategy for the mediator-outcome and exposure-outcome relationships. Fixed effects remove stable confounders but not time-varying ones, and observational mediation analysis depends on assumptions that no fixed-effects specification can supply (Imai, Keele and Tingley, 2010). A future empirical study could estimate the pathway legitimately only if it specified the mechanism in advance, estimated the components on a common sample, reported the indirect effect with a bootstrap interval that respects the clustering structure of the panel, and stated the confounding assumptions that the interval does not test. Until those conditions are met, the pathway should be described as a conceptual relationship to be investigated rather than as a mechanism.

Sensitivity, fit and the reading of a single statistic

The change in the AI coefficient between the one-period and two-period specifications, from 0.129 to -0.613, is the result most likely to be misread. Read alone, it suggests that the relationship is unstable. Read with the sample sizes attached to each estimate, it suggests that

two different quantities were estimated on two different sets of unit-period observations. The second reading is the correct one, and it depends on a comparison that the design did not perform. The full-model R-squared values of 0.720, 0.615 and 0.800 present a similar hazard. In a panel with forty unit effects and fifteen period effects, most of the variation in any outcome is absorbed by the fixed effects, and the resulting fit statistic measures the performance of the full model rather than the explanatory value of the regressors. Reporting an R-squared of 0.800 alongside a set of statistically inconclusive coefficients invites the reader to conclude that the model explains the outcome well, when what it demonstrates is that the fixed effects do their work. The two statistics are not in conflict; they answer different questions. The general lesson is that coefficients, intervals, sample sizes and fit statistics have to be read as a set. Any one of them, taken alone, can support a conclusion that the others contradict.

Implications for the empirical extension

Four design implications follow, and each is a condition that the empirical version of this study would have to meet in order to move from illustration to evidence. The lag structure must be derived from the mechanism rather than chosen for convenience. If the mechanism is that AI adoption changes production processes within a year or two, the specification should estimate that timing and should test alternatives on a common sample. If the mechanism is that invention precedes deployment by a longer interval, the specification should say so and should justify the interval. A lag that is neither derived nor justified tests nothing in particular. The AI measure must be described as what it is. Patent families record invention, and the protocol should pair the patent measure with at least one measure of adoption or diffusion, such as enterprise AI uptake, where coverage permits. Without that pairing, an association between invention and transformation is a statement about inventive activity rather than about the technology in use. The outcomes must be estimated separately and reported separately. Aggregation conceals the very trade-offs that motivate the framework, and a composite index whose weights are chosen after the estimates are known is a mechanism for producing the conclusion the researcher expected. The inference must respect the cluster structure. Forty clusters is the lower bound of defensibility, and a country-year panel over a fifteen-year period with a sample of forty countries would be at that bound. Where the number of clusters is smaller, wild-bootstrap procedures are the appropriate alternative (MacKinnon and Webb, 2017), and the reporting should state which procedure was used.

Limitations

The limitations are those of the design, stated in Section 2.9 and now restated in relation to the interpretation. The observations are constructed, so no conclusion about any economy follows. The variables are indices, so the exercise cannot speak to measurement error, which in patent and emissions data is substantial. A single data-generating process is examined, so the conclusions about the estimator are conditional on assumptions about persistence, heterogeneity and dependence that other processes would violate. The panel is balanced and complete, whereas cross-country panels are neither. The transformation construct used here is narrower than the process it represents. One further limitation is specific to the present reporting. The exact clustered standard errors for the two income specifications are not reported in the output from which Tables 3.2 and 3.3 were compiled, and the digital-capacity standard error in the baseline model is likewise not separately reported. These gaps do not affect the direction of any conclusion, because the intervals in Figures 3.1 and 3.2 show the same imprecision, but they would have to be closed before the paper were submitted for publication, and they are flagged in the table notes rather than left for a reader to discover. Finally, the paper does not examine heterogeneity. A single average coefficient across forty units describes no unit in particular, and the framework is explicit that digital capacity, human capital and institutional quality differ across settings. The absence of subgroup estimation is a consequence of the constructed panel, in which no such attributes exist, and it is the clearest indication of what the empirical extension would add.

Conclusion

Overview

This paper set out to examine the role of AI-driven innovation in green structural transformation and sustainable economic development, and to establish what the conventional panel design used in this literature can support as evidence. The framework proposed that AI-related capabilities may influence environmental and economic outcomes directly and through changes in productive activity, and identified digital capacity, human capital and institutions as conditions under which those effects are expected to operate. The framework was examined on a panel of forty economic units observed over fifteen periods, with unit and period fixed effects and standard errors clustered by unit. The analysis covered descriptive statistics, a baseline specification for green transformation, separate models for income, emissions intensity and employment, an interaction between AI innovation and digital capacity, and an alternative lag specification.

Principal conclusions

The analysis supports five conclusions, each conditional on the design described in Section 2.

1. The baseline association between one-period-lagged AI innovation and green transformation is positive but imprecise (0.129; standard error 0.405; $p = 0.750$). The estimate does not establish a positive relationship, and it does not establish the absence of one.
2. The income models return small negative coefficients (-0.027 and -0.029) at the 10 per cent level, with no attenuation when the transformation term is added. These estimates do not support a finding that AI innovation reduces income, and they do not support the proposed pathway.
3. The emissions and employment models identify no statistically distinguishable association, so the analysis cannot separate an environmental effect, an employment effect or their absence. The three development dimensions remain separate findings.
4. The interaction between AI innovation and digital capacity is negative and inconclusive (-1.184; $p = 0.184$), and it supports neither the anticipated complementarity nor its opposite. Institutional quality and human capital were not tested.
5. The AI coefficient changes sign between the one-period and two-period lag specifications, and the comparison confounds lag with sample. The reversal is not evidence of instability in a relationship; it is evidence that the specification does not identify a direction.

Contribution

The contribution is conceptual, methodological and procedural. Conceptually, the paper integrates AI innovation, green transformation and sustainable development into a single framework while keeping the three development dimensions separate, so that a favourable estimate on one dimension cannot be read as evidence about the others. Methodologically, it identifies four failure modes of a standard specification that are difficult to detect in observed data and easy to demonstrate when the generating process is known: a contemporaneous relationship estimated with a lag, an interaction interpreted without its constituent precision, a fit statistic inflated by fixed effects, and a change in sample composition mistaken for a change in lag structure. Procedurally, it specifies the indicator set and coverage constraints of an empirical extension before any estimate is reported, which prevents the variable definitions from being chosen after the results are known. A further contribution is the treatment of negative results. The paper reports five inconclusive estimates and one sign reversal, and it does not convert any of them into a claim about the world. That restraint is deliberate: the literature on AI and sustainable development contains a substantial number of findings that rest on marginal p-values, single-lag specifications and fit statistics that measure the wrong thing, and the remedy is not more analysis of the same kind but a clearer statement of what each statistic can support.

Research agenda

The next stage is to replace the constructed indices with verified observations and to align the specification with the mechanism it is meant to test. Six priorities follow from the analysis.

- Construct a documented country-year panel with consistent definitions and coverage, and report the country list, the years covered and the reasons for exclusion.
- Pair the patent-based measure of AI invention with at least one measure of AI adoption or diffusion, so that invention and use are not conflated.
- Derive the lag structure from the mechanism, estimate alternative timings on a common sample, and report the comparison rather than a single specification.
- Measure green transformation through multiple indicators rather than a single composite, and analyse the components separately before combining them.
- Test institutional, infrastructure and human-capital conditions separately, where coverage permits, rather than expecting a single digital indicator to capture complementarity.
- Estimate a pathway only when the temporal ordering matches the mechanism and the confounding assumptions can be stated, and report the indirect effect with a bootstrap interval that respects the clustering structure.

Closing

AI-driven innovation is a relevant subject for research on green structural transformation, but its contribution cannot be established from technological expectation alone. The literature identifies opportunities and risks in the same technology, which makes the conditions of implementation and governance the substance of the research question rather than a caveat attached to it. The present study does not determine whether AI delivers greener or more inclusive development. Its conclusion is that the evidential value of a panel design depends on four things that are usually decided casually: the timing of the relationship relative to the estimated lag, the correspondence between the measured construct and the mechanism it represents, the number of clusters available for inference, and the separation of outcomes that the framework treats as distinct. A defensible empirical paper on this question will be one in which those four decisions are made before the estimates are seen and reported alongside them.

References

[1] T. Babina, A. Fedyk, A. He, and J. Hodson, “Artificial intelligence, firm growth, and product innovation,” *Journal of Financial Economics*, vol. 151, Art. no. 103745, 2024.

[2] A. C. Cameron and D. L. Miller, “A practitioner's guide to cluster-robust inference,” *Journal of Human Resources*, vol. 50, no. 2, pp. 317–372, 2015.

[3] K. Imai, L. Keele, and D. Tingley, “A general approach to causal mediation analysis,” *Psychological Methods*, vol. 15, no. 4, pp. 309–334, 2010.

[4] J. G. MacKinnon and M. D. Webb, “Wild bootstrap inference for wildly different cluster sizes,” *Journal of Applied Econometrics*, vol. 32, no. 2, pp. 233–254, 2017.

[5] Organisation for Economic Co-operation and Development (OECD), *Identifying Emerging AI Technologies Using Patent Data*. Paris, France: OECD Publishing, 2025.

[6] OECD, “Patents by technology,” *OECD Data Explorer*, [Online]. Available: OECD Data Explorer. [Accessed: Nov. 24, 2026].

[7] OECD, “Patents on environment technologies,” *OECD Data Explorer*, [Online]. Available: OECD Data Explorer. [Accessed: Nov. 24, 2026].

[8] F. Amin, “Binary flaw detection: A security analysis paper,” in *Proc. 2025 Int. Conf. Advances in Machine Intelligence and Cybersecurity Technologies (AMICT)*, Kota Kinabalu, Sabah, Malaysia, 2025, pp. 325–330, doi: 10.1109/AMICT65811.2025.11402666.

[9] F. Amin, “A scalable framework for interpretable binary vulnerability analysis using data dependency modeling,” in *Proc. Int. Conf. Artificial Intelligence and Cybersecurity (ICAIC)*, 2025, pp. 249–255.

[10] A. Khan, U. Imtiaz, and F. Amin, "Honeytoken-augmented AI model for insider threat detection using cyber deception and isolation forest," *Kashf Journal of Multidisciplinary Research*, vol. 2, no. 9, pp. 294–313, 2025.
[11] N. A. Al-Sultani, "An advanced AI and machine learning framework for IoT-driven data science, edge intelligence, real-time pattern recognition, and autonomous decision making," *Journal of Intelligent Decision Making and Information Science*, vol. 3, no. 17s, pp. 970–998, 2026.
[12] V. Gummadavelli, U. Imtiaz, P. A, S. H. Kolla, N. S. Bavirisetty, and S. R. Kondapalli, "Discovering and repairing blind spots in LLM-as-a-judge evaluation of multi-turn AI systems," in *Proc. 2026 Int. Conf. Intelligent and Sustainable AI Systems (ICOSAAS)*, Bali, Indonesia, 2026, pp. 1611–1618, doi: 10.1109/ICOSAAS68663.2026.11648543.
[13] I. U. Rahman, F. Ijaz, A. Kalsoom, M. Altaf, and S. B. Batool, "The green structural transformation: Climate change, industrial policy, and development in a fragmenting global economy," *Inverge Journal of Social Sciences*, vol. 5, no. 5, pp. 43–65, 2026, doi: 10.63544/ijss.v5i5.335.
[14] United Nations Conference on Trade and Development (UNCTAD), *Technology and Innovation Report 2025: Inclusive Artificial Intelligence for Development*, UNCTAD/TIR/2025. United Nations, 2025.
[15] R. Vinuesa *et al*., "The role of artificial intelligence in achieving the Sustainable Development Goals," *Nature Communications*, vol. 11, Art. no. 233, 2020, doi: 10.1038/s41467-019-14108-y.
[16] World Bank, "Manufacturing, value added (% of GDP)," *World Development Indicators*, [Online]. Available: World Development Indicators. [Accessed: Nov. 24, 2026].
[17] World Bank, *World Development Indicators*, [Online]. Available: https://databank.worldbank.org/source/world-development-indicators.
[18] World Bank, *Worldwide Governance Indicators*, [Online]. Available: https://www.worldbank.org/en/publication/worldwide-governance-indicators.
[19] B. Behera, P. Behera, U. K. Pata, L. Sethi, and N. Sethi, "Artificial intelligence-driven green innovation for sustainable development: Empirical insights from India's renewable energy transition," *Journal of Environmental Management*, vol. 389, Art. no. 126285, 2025, doi: 10.1016/j.jenvman.2025.126285. PubMed
[20] Q. Liu, S. Du, and M. Li, "Green innovation perspective: Artificial intelligence and corporate green development," *International Review of Economics & Finance*, vol. 102, Art. no. 104345, 2025, doi: 10.1016/j.iref.2025.104345. ScienceDirect
[21] B. Lin and D. Zhou, "A new green transition driver: How does artificial intelligence affect the green economic efficiency," *Energy*, vol. 334, Art. no. 137808, 2025, doi: 10.1016/j.energy.2025.137808. ScienceDirect
[22] M. Skare, B. Gavurova, and D. Sinkovic, "Measuring artificial intelligence's impact on sustainable energy transition: Empirical insights and policy implications," *Energy Economics*, vol. 150, Art. no. 108825, 2025, doi: 10.1016/j.eneco.2025.108825. ScienceDirect
[23] X. Wang, "How does artificial intelligence accelerate the energy transition? Learning from empirical experience in OECD countries," *Journal of Environmental Management*, vol. 391, Art. no. 126397, 2025, doi: 10.1016/j.jenvman.2025.126397. PubMed
[24] H. Li, "Artificial intelligence, green transition and green total factor productivity in enterprises," *Scientific Reports*, vol. 15, Art. no. 33248, 2025, doi: 10.1038/s41598-025-18878-y. Nature
[25] J. Chen, L. Yang, and H. Su, "Green transformation in the era of intelligence: How artificial intelligence affects disruptive green innovation in firms," *Energy Economics*, vol. 153, Art. no. 109093, 2026, doi: 10.1016/j.eneco.2025.109093