Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Renewable Capacity, Coal Spillovers, and Grid Reliability: Spatial Durbin and LeSage-Pace Panel Evidence from Indian States Under SDG 7

Prakash Gurumurti Vhankade

prakash.vhankade@gipe.ac.in

Gokhale Institute of Politics and Economics https://orcid.org/0009-0005-2436-0603

Research Article

Keywords: SDG 7, energy reliability, capacity-generation gap, panel threshold regression, Spatial Durbin Model, Driscoll-Kraay standard errors, India

Posted Date: July 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10395391/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Renewable Capacity, Coal Spillovers, and Grid Reliability: Spatial Durbin and LeSage-Pace Panel Evidence from Indian States Under SDG 7

Dr. Prakash Gurumurti Vhankade
Gokhale Institute of Politics and Economics, Pune, Maharashtra, India
prakash.vhankade@gipe.ac.in
ORCID: 0009-0005-2436-0603

Abstract

Sustainable Development Goal 7 commits countries to affordable, reliable, sustainable, and modern energy, in which reliable is a distinct operational target that connection alone does not satisfy. India has achieved near-universal household electrification and reduced its national energy shortage to 0.1 percent in 2024-25, yet installed renewable capacity represents 48 percent of its total installed base while coal accounts for roughly 71 percent of actual generation, revealing a persistent capacity-generation gap. This paper tests whether that gap has a measurable counterpart in state-level reliability outcomes. A balanced panel of 35 Indian states from 2020 to 2023 assembled from sixteen Central Electricity Authority and Ministry datasets supports six empirical analyses. A two-way fixed effects model with Driscoll-Kraay standard errors, motivated by strong cross-sectional dependence (Pesaran CD: $z = 4.493$, $p < 0.001$), finds coal capacity share negatively and significantly associated with state-level energy deficit ($p = 0.002$); renewable capacity share shows no detectable association across twelve estimation environments. A Goodman-Bacon decomposition reveals a between-versus-within sign reversal for coal capacity. A Mundlak-Chamberlain Wald test confirms fixed and random effects are statistically equivalent. A Hansen threshold test finds no interior regime shift. A Spatial Durbin Model with LeSage-Pace impacts decomposition finds significant positive indirect effects of neighbouring states' coal capacity on own-state deficit (0.279, $p = 0.010$; 0.178, $p = 0.024$), while Moran's I tests confirm no spatial autocorrelation in the outcome. All findings are correlational. The results argue that capacity-based SDG 7.2 monitoring overstates the operational maturity of the transition.

Keywords: SDG 7; energy reliability; capacity-generation gap; panel threshold regression; Spatial Durbin Model; Driscoll-Kraay standard errors; India

1. Introduction

Sustainable Development Goal 7 sets a specific and multidimensional standard for energy progress: access to energy that is affordable, reliable, sustainable, and modern, for all. The word reliable is not decorative. A household connected to the grid but receiving supply that falls short of demand in aggregate has not received what the goal describes. The monitoring literature has tracked access comprehensively through indicator 7.1.1 and renewable capacity through indicator 7.2.1, but has not disaggregated the quality of supply once connection is achieved to the subnational level where procurement, scheduling, and distribution decisions are actually made.

India's energy sector presents the question with exceptional clarity. The national electrification record is unambiguous: the Saubhagya programme connected virtually all remaining unelectrified households, and India's aggregate energy shortage declined from 4.2 percent of total energy requirement in 2013-14 to 0.1 percent in 2024-25 (Ministry of Power, Government of India, 2025). The renewable transition record is equally prominent: installed renewable capacity reached 220 gigawatts by 2024-25, representing 48 percent of total installed capacity, with solar and wind together at 156 gigawatts (Ministry of Power, Government of India, 2025). India was among only ten countries globally on track to triple non-fossil generation capacity from 2022 levels by 2030 (Ember, 2025). These milestones constitute genuine progress on the capacity dimension of SDG 7.

A harder question lies beneath these milestones. Installed capacity and dispatched generation are distinct quantities. A generating unit installed but idle because of intermittency, merit-order displacement, grid congestion, or minimum technical load constraints does not contribute to supply adequacy at peak demand hours. Coal accounted for approximately 71 percent of India's electricity generation in 2024 despite representing 47 percent of installed capacity, while solar and wind accounted for approximately 12 percent of generation despite representing 33 percent of installed capacity (Ministry of Power, Government of India, 2025; Ember, 2025). This capacity-generation divergence is the structural condition this paper investigates at the state level, using the energy deficit percentage, the shortfall of actual electricity availability relative to stated requirement, as the primary measure of whether supply adequacy is being achieved.

The Tracking SDG 7: The Energy Progress Report, produced annually by the IEA, IRENA, UNSD, World Bank, and WHO, is the authoritative reference for monitoring Goal 7's sub-targets (IEA et al., 2024; IEA et al., 2025). The report monitors India's progress on indicator 7.2.1 primarily through installed renewable capacity share. Whether that capacity share is converting into reliable dispatched electricity at the state level, and whether it is associated with measurable reductions in state-level energy deficits, are questions the national indicator cannot answer. Manojkumar et al. (2026) document that most of India's coal fleet operates at minimum technical loads of approximately 55 percent under binding power purchase agreements regardless of renewable availability, maintaining coal's effective dispatch-order priority. Li et al. (2025) document an

equivalent capacity-generation mismatch in Chinese provinces driven by transmission and storage constraints. Lu et al. (2022) show for China that expanding grid balancing areas, not coal plant flexibility retrofits, is the effective remedy. This paper provides the first state-level panel test of whether these structural features produce a detectable capacity-generation gap in India's reliability data.

The empirical strategy proceeds through six stages. A two-way fixed effects model with Driscoll-Kraay standard errors (Driscoll and Kraay, 1998) corrects for strong cross-sectional dependence confirmed by the Pesaran (2004) CD test ($z = 4.493$, $p < 0.001$, mean absolute pairwise residual correlation = 0.489). A Goodman-Bacon (2021) decomposition characterises the composition of the two-way FE coefficient, revealing a between-versus-within sign reversal that restricts interpretation to within-state temporal associations. A Mundlak-Chamberlain correlated random effects specification (Mundlak, 1978; Chamberlain, 1982) provides a formal Wald test of the Hausman endogeneity concern, failing to reject the equivalence of fixed and random effects (chi-squared = 3.797, $p = 0.150$). A Hansen (1999) panel threshold model tests for nonlinear regime shifts in the renewable-reliability association. A Spatial Durbin Model (Elhorst, 2014), estimated under a grid-region weight matrix grounded in India's five official regional grid zones and a geographic contiguity matrix as robustness, provides LeSage-Pace (2009) direct, indirect, and total impacts decomposition with Monte Carlo standard errors. Five targeted robustness checks test stability across alternative reliability measures, subsamples, quantiles, relative importance decomposition, and specification of individual effects.

The paper produces four principal findings. Coal capacity share is negatively and significantly associated with state-level energy deficit in the primary two-way FE model under Driscoll-Kraay standard errors (coefficient -0.102, $p = 0.002$), an association that strengthens when cross-sectional dependence is corrected relative to the conventional clustered result ($p = 0.035$). Renewable capacity share shows no detectable association with energy deficit across twelve estimation environments, a result that recurs across estimators, spatial specifications, and quantiles with sufficient consistency to constitute a structural finding rather than a statistical artefact. The LeSage-Pace impacts decomposition reveals that coal capacity generates a significant positive indirect effect on own-state deficit through neighbouring states' capacity composition, robust to grid-region weighting (0.279, $p = 0.010$) and contiguity weighting (0.178, $p = 0.024$), while Moran's I tests confirm no direct spatial autocorrelation in the outcome. The Goodman-Bacon decomposition exposes a between-versus-within sign reversal: persistently coal-heavy states show higher average deficits cross-sectionally, while within each state, years of higher coal share show lower deficits. All findings are correlational; the observational strategy does not support causal claims.

Three implications follow for the SDG 7 monitoring literature. Installed renewable capacity share, the primary instrument for gauging progress on indicator 7.2.1 in the Tracking SDG 7 framework, is not converting into a measurable reliability contribution at the state level during 2020-2023.

Monitoring SDG 7.2 progress through capacity share alone may overstate the operational maturity of the transition. State-level generation share measures whether installed renewable capacity is actually dispatched; it is a closer indicator of whether SDG 7.1's reliability dimension is being served. The spatial coal spillover finding indicates that the inter-state reliability buffer currently embedded in India's regional grid architecture operates through coal-based dispatch; renewable capacity has not yet acquired a comparable spillover function.

The remainder of the paper is organised as follows. Section 2 reviews the relevant literature and develops the theoretical framework. Section 3 describes the data, variable construction, and estimation strategy. Section 4 presents the empirical results. Section 5 discusses the findings in relation to the SDG 7 framework and the capacity-generation gap literature. Section 6 draws policy implications. Section 7 concludes.

2. Review of Literature and Theoretical Framework

2.1 SDG 7 Progress Tracking and the Reliability Dimension

The annual Tracking SDG 7: The Energy Progress Report, produced jointly by the International Energy Agency, the International Renewable Energy Agency, the United Nations Statistics Division, the World Bank, and the World Health Organization, is the authoritative global reference for monitoring progress against Goal 7's four sub-targets. The 2024 edition documents that despite measurable improvements across indicators, the current pace is insufficient to achieve any of the 2030 targets; the 2025 edition reports that global electricity access reached 92 percent in 2023, leaving 666 million people without basic access (IEA et al., 2024; IEA et al., 2025). These reports track indicator 7.1.1, indicator 7.2.1, and indicator 7.3.1 at country level and do not disaggregate the quality of supply once households are connected. This subnational reliability gap motivates the state-level panel approach adopted in the present paper.

The distinction between electricity access and electricity reliability is not merely terminological. SDG 7's stated objective is access to affordable, reliable, sustainable, and modern energy for all, in which reliable is a specific qualifier that connection alone does not satisfy. A household connected to the grid but receiving supply that falls short of demand does not have what the goal describes. India's reduction of aggregate energy shortages from 4.2 percent in 2013-14 to 0.1 percent in 2024-25 at the national level reflects genuine progress on this reliability dimension (Ministry of Power, Government of India, 2025). The persistence of state-level deficits during the intervening sample period 2020-2023, documented in this paper, indicates that national improvement conceals significant subnational heterogeneity.

Applied econometric work has begun supplementing indicator-level tracking with subnational spatial analysis. Azaliah et al. (2023), published in this journal's SDG 7 article collection, test energy intensity convergence across 33 Indonesian provinces from 2010 to 2018 using a spatial

Durbin framework grounded in Elhorst (2014), finding evidence of conditional beta convergence without sigma convergence. That study establishes a direct methodological precedent for applying spatial panel econometrics to energy outcomes within a single developing country. The precedent is particularly pertinent because it was published in the same journal and SDG 7 collection, confirming the acceptance of this methodological strand for this type of subnational question.

2.2 The Capacity-Generation Gap: Theory and International Evidence

The divergence between installed renewable capacity and realized renewable generation is a structural feature of energy transitions in large federal power systems. Its origins lie in four mechanisms: the intermittency of solar and wind output, the position of variable renewable generation behind dispatchable thermal units in merit-order dispatch, the absence of grid-scale storage infrastructure to absorb off-peak renewable surpluses, and transmission congestion that prevents the transport of surplus renewable generation from resource-rich to demand-heavy regions. Each mechanism independently constrains the conversion of installed capacity into dispatched electricity, and they interact: congestion that prevents inter-regional transfer compounds the effect of intermittency during peak demand hours when renewable output is low.

The IEA (2024) documents that maintaining grid stability in the energy transition requires transmission networks to absorb a more distributed and variable generation mix; over 3,000 gigawatts of renewable capacity awaited grid connection permits globally in 2023, reflecting a structural lag between capacity installation and grid integration. This global pattern has its closest documented parallel in China, where Li et al. (2025) document pronounced spatiotemporal mismatch between renewable energy production and consumption across Chinese provinces, arising from inadequate inter-regional transmission, limited storage, and geographic misalignment between renewable resource endowment and industrial demand. Lu et al. (2022) demonstrate for China that expanding balancing areas from the provincial to the national level reduces curtailment significantly, while retrofitting coal plants for greater operational flexibility offers comparatively limited improvement. This finding has direct implications for India's regional grid architecture: if the mechanism limiting renewable integration is grid span rather than thermal inflexibility, transmission expansion rather than plant retrofits is the priority policy instrument.

In Germany, Shu et al. (2020) identify grid congestion in the transmission system as the primary source of renewable curtailment, with curtailment payments rising from 43.7 million euros in 2013 to 718.7 million euros in 2018 as wind penetration increased without commensurate transmission expansion. The REN21 (2024) Global Status Report corroborates these patterns globally, citing Vietnam in 2023 as an example where network congestion left over half of installed solar capacity unused at peak times and Japan where curtailment reached a record high in 2023-24, directly linking grid infrastructure bottlenecks to the capacity-generation gap. In India, Manojkumar et al. (2026) document that most coal plants operate at minimum technical loads of approximately 55 percent

under binding long-term power purchase agreements, keeping thermal output continuous regardless of renewable availability. This operational rigidity maintains coal's role as the effective baseload and reliability anchor, independent of its cost competitiveness relative to renewable alternatives, and is the structural mechanism that the present paper's panel results reflect.

Davidson et al. (2016) identify fragmented dispatch authority and long-term thermal contracts as institutional drivers of curtailment that persist even as renewable capacity expands in China, a finding that applies directly to India's comparable regulatory structure of state-level power purchase agreements and central and state sector generation assets with different dispatch priorities.

2.3 Discom Financial Health and the Omitted Variable Problem

A well-established body of applied work on the Indian power sector has identified the financial and operational condition of electricity distribution companies as the primary determinant of supply adequacy at the state level, independent of the generation capacity mix. This body of work defines the dominant omitted variable category in the present paper's specification and directly informs the interpretation of the low within-R-squared.

Verma et al. (2020) provide a comprehensive critical review of Indian power distribution sector reforms, documenting that most state-owned discoms carry severe financial stress from high aggregate technical and commercial losses, unpaid electricity bills, outstanding regulatory assets, and aging distribution infrastructure, all of which constrain their ability to procure and schedule power reliably. Singh et al. (2024) extend this analysis with panel data on institutional reforms across state utilities, finding that generation, transmission, and distribution capacities remain misaligned with demand growth and identifying the distribution segment as the weakest link in the supply chain. The Power Finance Corporation's annual performance report records that national AT&C losses rose from 15.11 percent in 2022-23 to 16.12 percent in 2023-24, with state-level variation spanning below 2 percent in best-performing utilities to above 30 percent in chronically stressed ones (Power Finance Corporation, 2024). The Revamped Distribution Sector Scheme and its predecessor UDAY reflect central government recognition that discom financial health is the binding constraint on reliability in chronically deficit states, not generation capacity per se (Ministry of Power, Government of India, 2024).

These findings present a specific omitted variable challenge for the present paper. States with better discom financial health tend both to maintain higher coal capacity shares through better-financed procurement and to show lower energy deficits through better supply scheduling and lower AT&C losses. The two-way FE estimator controls for the time-invariant component of discom financial health. The Mundlak-Chamberlain CRE specification further tests whether the time-varying component is correlated with the capacity composition regressors. Both address the endogeneity concern partially; neither resolves it completely, a limitation stated explicitly in the inference and scope statements throughout.

2.4 Coal and Renewable Energy in India's Power Sector

Mishra et al. (2023), applying Johansen cointegration, regime-shift cointegration, and nonlinear autoregressive distributed lag estimation to Indian coal production, industrial output, and electricity generation data from 1999 to 2020, project continued coal dependence for the majority of India's electricity demand through the mid-2020s notwithstanding accelerating renewable deployment. This projection is consistent with the observed data: as of December 2025, India had achieved 258 gigawatts of installed renewable capacity but coal continued to generate approximately 71 percent of electricity in 2024, with renewables capturing a larger share of incremental demand in 2025 only under unusually low-demand conditions (Ember, 2025).

Gulagi et al. (2022), using the LUT Energy System Transition model, find that a fully renewables-based power system across Indian states by 2050 would be cost competitive with the current coal-dominated system while maintaining reliability, and identify inter-state transmission expansion as a necessary condition for smoothing renewable resource variability across states. The Gulagi et al. study is a forward-looking capacity optimization model, not an estimation of realized historical outcomes; its conclusions rest on storage and transmission buildout that has not materialised during the 2020-2023 period examined in this paper. The contrast between Gulagi et al.'s long-run scenario and the present paper's short-run panel findings is substantive rather than contradictory: the conditions under which renewable generation reliably substitutes for coal dispatch are projected but not yet achieved.

2.5 Theoretical Framework: Reliability, Dispatchability, and Spatial Interdependence

The paper's empirical strategy rests on a theoretical framework that distinguishes installed capacity, dispatchable capacity, and delivered supply adequacy, and treats state-level reliability as spatially interdependent through the regional grid architecture.

Installed capacity is a stock variable measuring maximum possible generation output under ideal conditions. Dispatchable capacity is the subset of installed capacity that can be scheduled, ramped, and delivered to meet demand at specific hours, including peak demand periods that drive aggregate energy deficits. The capacity-generation gap arises because installed renewable capacity converts to dispatchable output at a lower effective rate than thermal capacity during the current transition period, due to intermittency, dispatch-order constraints, and curtailment. The relevant theoretical prediction is that renewable capacity share may not carry a detectable direct association with aggregate energy deficit during a transition period in which the conversion efficiency of renewable capacity into dispatched generation remains low, the capacity-generation gap is open, and complementary storage and transmission infrastructure has not yet closed it.

Coal capacity's continuous dispatch at or above minimum technical load gives it a higher effective conversion rate from installed capacity to delivered supply. The theoretical prediction is

accordingly that coal capacity share carries a negative association with energy deficit conditional on state and year effects, since states with a larger coal share have access to more dispatchable generation per unit of installed capacity.

The spatial interdependence of state-level reliability operates through India's five regional grid zones, within which generating units are pooled and surplus power from one state can be dispatched to address shortfall in another through inter-state transmission. This architecture generates a prediction for spatial spillovers in capacity composition: the coal capacity share of grid-zone neighbours may affect own-state deficit by altering the availability of dispatchable surplus within the zone. The sign of this indirect effect is ambiguous a priori. If neighbours' coal capacity primarily generates exportable surplus, the indirect effect should be negative: more neighbour coal capacity, lower own deficit. If neighbours' coal capacity is primarily consumed internally or if coal-heavy states cluster in high-demand zones, the indirect effect may be positive. The Spatial Durbin Model is the appropriate specification for this test because it separately identifies the direct association of own-state coal capacity and the indirect association of neighbours' coal capacity through the LeSage-Pace impacts decomposition.

2.6 Econometric Framework: Panel Methods, Heterogeneity, and Spatial Dependence

The econometric methods used in this paper draw on four distinct strands of the panel data and spatial econometrics literature, each engaging a specific methodological concern raised by the data structure.

Two-way fixed effects and treatment effect heterogeneity. The two-way fixed effects estimator removes time-invariant state-level confounders and common year-level shocks from panel estimates. Goodman-Bacon (2021) demonstrates that in panels with variation in treatment intensity across units and time, the TWFE coefficient is a weighted average of all possible 2x2 difference-in-differences comparisons, with weights based on subsample shares and variance in treatment timing. When treatment effects are heterogeneous, some weights may reflect comparisons in which already-treated units serve as controls for later-treated units, producing coefficients that do not recover a convex combination of average treatment effects. Callaway and Sant'Anna (2021) develop a heterogeneity-robust alternative restricting comparisons to clean groups, but its application requires a binary treatment variable and sufficient time periods for group-time average treatment effects. For the continuous treatment variables and short time dimension ($T = 4$) in this paper, the Callaway-Sant'Anna estimator is not feasible. The Goodman-Bacon decomposition is applied as a diagnostic to characterise the composition of the TWFE estimate and the scope restriction that follows from it.

Correlated random effects. The Mundlak (1978) and Chamberlain (1982) correlated random effects framework tests the Hausman (1978) endogeneity concern without requiring the full fixed effects specification. Augmenting a random effects model with the within-state time-means of all

time-varying regressors, the CRE model tests whether state-level random effects are correlated with the regressors. A Wald test of the joint null that the correction terms equal zero provides formal evidence on the equivalence of FE and RE, resolving the Hausman test ambiguity that arises when T is too small to support two-way random effects estimation.

Cross-sectional dependence and robust inference. Pesaran (2004) provides a panel test for cross-sectional dependence based on the average pairwise correlation of model residuals, consistent under both weak and strong cross-sectional dependence. Where cross-sectional dependence is detected, Driscoll and Kraay (1998) provide an asymptotically valid covariance matrix estimator robust to heteroskedasticity, serial correlation, and cross-sectional dependence of unknown form, implemented as the spatial and serial correlation robust estimator. Driscoll-Kraay standard errors are preferred over state-clustered standard errors when cross-sectional dependence is strong, as confirmed in this paper's Pesaran CD test.

Spatial Durbin Model and impacts decomposition. Elhorst (2014) provides the foundational treatment of spatial panel models, including the specification and interpretation of the Spatial Durbin Model. LeSage and Pace (2009) establish that raw SDM coefficients are not marginal effects because they conflate direct associations with the spatial multiplier process induced by the spatial autoregressive parameter. The correct decomposition into average direct, indirect, and total impacts proceeds through the impacts matrix S(W) = (I - lambdaW)^{-1}, with Monte Carlo simulation used to obtain standard errors. This decomposition is essential for interpreting the spatial coal spillover finding reported in this paper.

2.7 Positioning of the Present Study

Five gaps in the existing literature motivate this paper's contribution.

First, the Tracking SDG 7 framework monitors reliability-adjacent indicators at country level but does not disaggregate supply quality to the subnational level. State-level panel evidence on the reliability dimension of SDG 7.1 is absent from the monitoring literature.

Second, while the capacity-generation gap is documented at the national level for India, it has not been tested econometrically as an explanation for the absence of a measurable reliability benefit from renewable capacity in state-level panel data.

Third, the discom financial health literature has established AT&C losses and ACS-ARR gaps as primary determinants of supply adequacy, but this literature has not been integrated with capacity-composition analysis in a spatial panel framework that separates own-state and cross-state effects.

Fourth, the Spatial Durbin Model has been applied to cross-country renewable energy spillovers but rarely to within-country, state-level grid reliability outcomes in a federal power system organised around defined regional grid zones.

Fifth, the recent literature on TWFE heterogeneity (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021) has not been applied to subnational energy reliability studies in India, where the

continuous treatment and short panel dimension present specific constraints on heterogeneity-robust estimation that warrant explicit engagement rather than silence.

3. Data and Methods

3.1 Data Sources

The analysis draws on sixteen state and national-level datasets covering India's power sector, sourced from central government and public sector energy reporting agencies and publicly available through India's open data portal and the Central Electricity Authority. Table A1 in the Appendix lists each dataset with its originating agency, temporal coverage, observation frequency, and unit of analysis. The datasets collectively span daily power generation records (2017-2025), daily power outage records (2017-2025), state-level renewable energy generation records (2019/2020-2025 depending on series), monthly energy requirement and availability records (January 2020 to June 2023), installed capacity by fuel type and state (2017-2025), state-level coal consumption (2006-2021), and national transmission line commissioning records (2014-2025). Coverage windows differ across datasets; the estimation sample described in Section 3.3 is determined by their intersection.

3.2 Variable Construction

All daily and monthly source series are aggregated to the state-year level. Six variables are constructed.

Energy deficit percentage is the percentage shortfall of annual energy availability relative to annual energy requirement, summed from monthly state-level records and expressed as a positive number when availability falls short of requirement. This is the primary outcome variable.

Outage rate is the share of generating-unit station-days on which an active outage of any type is recorded, aggregated from daily unit-level records. This is the alternative reliability outcome used in robustness check 1, and is conceptually distinct from energy deficit: deficit captures system-level aggregate shortfall governed by capacity and dispatch structure, while outage rate captures unit-level operational disruption governed by maintenance and fuel supply factors. The two reliability measures are not interchangeable, and findings specific to one are explicitly scoped to that measure throughout.

Renewable generation share is the ratio of total renewable energy generated to total electricity generated from all sources, both aggregated from daily records. This variable serves as the threshold variable in the Hansen nonlinearity test, because it reflects the actual penetration of renewable output into the grid rather than installed capacity, which may be idle or curtailed.

Renewable capacity share and coal capacity share are the average annual shares of installed renewable and coal capacity, respectively, in total installed capacity across all fuel types. These are the primary explanatory variables in the fixed effects and spatial specifications.

Coal consumption is the annual state-level total from the coal consumption dataset. It is available only through 2021 and is absent for states with no domestic coal consumption. It is excluded from the primary specification and used only in robustness check 2.

State names are harmonized across all sixteen source files before merging. Corrections address inconsistent spellings, the 2020 administrative merger of Dadra and Nagar Haveli with Daman and Diu, and the October 2019 bifurcation of Jammu and Kashmir into the union territories of Jammu and Kashmir and Ladakh. The two post-bifurcation units are treated as a single combined unit across the full panel to preserve balance, at the cost of not separately identifying their outcomes after 2019. Bhutan, which appears in Indian power sector records due to cross-border hydropower trade, and national aggregate rows are excluded.

3.3 Panel Construction and Estimation Sample

The merged state-year panel is restricted to 2020 through 2023, the period over which energy requirement and availability data and state-level renewable generation data are both populated. This yields a balanced panel of 35 states and union territories over four years, N = 140 state-year observations. Renewable generation share values recorded as non-finite, arising from division by zero in states with negligible total generation, are recoded to zero. The underlying condition, an absence of renewable generation activity, is substantive information rather than a data gap. Cumulative national transmission line length varies by year and not by state; it is absorbed entirely by year fixed effects in the two-way model and does not appear as a separately identified coefficient.

3.4 Estimation Strategy

3.4.1 Primary Panel Model

The primary specification is a two-way fixed effects model:

$$Yit = ai + gt + b1\, res_cap_shareit + b2\, coal_cap_shareit + eit$$

where ai are state fixed effects, gt are year fixed effects, and eit is the idiosyncratic error. The within estimator is used in preference to the between or random effects estimator because it is consistent under any distributional assumption on ai, including the case where the individual effects are correlated with the regressors (Wooldridge, 2010). This consistency is unconditional and does not depend on the validity of any additional assumption.

Standard errors are reported under two corrections applied sequentially. The primary correction clusters standard errors at the state level, which accounts for within-state serial correlation and heteroskedasticity. The secondary correction applies Driscoll-Kraay (1998) standard errors, which

are robust to cross-sectional dependence, heteroskedasticity, and serial correlation simultaneously, implemented via the vcovSCC estimator. The Driscoll-Kraay correction is the preferred basis for inference: the Pesaran (2004) CD test strongly rejects the null of no cross-sectional dependence in the panel residuals ($z = 4.493$, $p < 0.001$), with a mean absolute pairwise residual correlation of 0.489 across all 595 state pairs, confirming that state-level clustering alone is insufficient. Both sets of standard errors are reported in full.

The choice between fixed and random effects is assessed through three formal procedures. First, an F-test of the joint significance of state and year fixed effects against pooled OLS rejects the pooled specification ($F = 3.31$, $p < 0.001$). Second, a Hausman test comparing individual fixed effects against random effects fails to reject consistency of the random effects estimator (chi-squared = 2.16, df = 2, p = 0.340). Third, to resolve this tension formally, a Mundlak-Chamberlain Correlated Random Effects (CRE) model is estimated, augmenting the random effects specification with the state-level time-means of all time-varying regressors as additional controls (Mundlak, 1978; Chamberlain, 1982):

$$Yit = m + b1\, res_cap_shareit + b2\, coal_cap_shareit + d1\, res - bari + d2\, coal - bari + ui + eit$$

where res-bari and coal-bari are the within-state time-means of the two regressors and ui is the random individual effect. A Wald test of the joint null H0: d1 = d2 = 0 returns chi-squared = 3.797 (df = 2, p = 0.150), failing to reject the null. The CRE model confirms formally that the individual effects are not systematically correlated with the regressors in this panel. The two-way FE model is retained as the primary specification for unconditional consistency and continuity with the spatial models. The Hausman test result is characterised as a potential efficiency cost rather than a specification error.

A further consideration concerns heterogeneous treatment effects in the TWFE estimator. Goodman-Bacon (2021) demonstrates that two-way FE coefficients in panels with variation in treatment intensity across units and time are weighted averages of all possible 2x2 difference-in-differences comparisons, with weights proportional to group size and variance in treatment status. A manual Goodman-Bacon decomposition, applied to a median-split binary proxy for coal capacity share to enable 2x2 pairwise comparisons, identifies 306 state-pair comparisons. The weighted average of these comparisons is +0.601, opposite in sign to the within-state FE coefficient of -0.102. This sign reversal reflects a between-versus-within divergence: states that are persistently high-coal tend to have higher average deficits than persistently low-coal states in cross-sectional comparisons, while within each state, years of higher coal share are associated with lower deficits. The TWFE coefficient captures the within-state temporal association, not the cross-sectional ranking of states by coal capacity. This scope restriction is maintained throughout all interpretation. Of the 306 state-pair comparisons, 130 (42.5 percent) produce negative 2x2 estimates, indicating substantial

treatment effect heterogeneity across state pairs. In this setting, the Callaway-Sant'Anna (2021) group-time average treatment effects estimator would provide a heterogeneity-robust alternative; however, the continuous treatment variable and short panel (T = 4) prevent its application here. All coefficients are interpreted as variance-weighted averages of within-state associations rather than as homogeneous causal parameters. All findings in this paper are correlational; the observational identification strategy does not support causal claims.

3.4.2 Cross-Sectional Dependence and Standard Error Choice

Prior to spatial modelling, the presence and structure of cross-sectional dependence in the panel residuals is assessed formally through two tests. The Pesaran (2004) CD test on the primary FE model residuals rejects the null of no cross-sectional dependence ($z = 4.493$, $p < 0.001$). The Frees (1995) test returns an average pairwise residual correlation of rho = 0.092. The mean absolute pairwise residual correlation computed directly across all 595 state pairs is 0.489. This level of contemporaneous cross-state error correlation is consistent with common shocks to the Indian electricity grid, including national coal supply disruptions, all-India aggregate demand surges, and monsoon-driven hydropower output fluctuations, that affect multiple states simultaneously regardless of their individual capacity composition.

Two implications follow directly. State-level clustered standard errors are insufficient because they do not correct for the contemporaneous cross-state correlation documented above; Driscoll-Kraay standard errors are therefore the preferred correction for inference throughout this paper. The presence of strong cross-sectional dependence in residuals that is not absorbed by state or year fixed effects also constitutes prior structural evidence that a spatial modelling framework is appropriate for characterising the inter-state dependence structure in this data.

3.4.3 Threshold Nonlinearity

A Hansen (1999) panel threshold model tests whether the associations between capacity composition and energy deficit change discontinuously beyond a critical level of renewable generation share. Renewable generation share is used as the threshold variable because it reflects actual grid penetration of renewable output rather than installed capacity, which may be idle or curtailed. The threshold value is estimated by a grid search over the 5th to 95th percentile of the threshold variable's distribution, minimising the residual sum of squares across all candidate threshold values. Statistical significance is assessed by the bootstrap F-test of the null hypothesis of no threshold effect, using 300 replications.

A critical condition for reliable threshold estimation is that the observed data contain sufficient variation to support a genuine interior break point. In panels with a short time dimension or limited cross-sectional variation in the threshold variable, the grid search may converge to the boundary of the search range rather than to an interior inflection point. Such a boundary-degenerate result should

be interpreted as a genuine null rather than as a specification failure, consistent with the inferential conditions stated in Hansen (1999). Both the point estimate and the bootstrap p-value are reported; a boundary-degenerate estimate accompanied by a bootstrap p-value above conventional significance thresholds is treated as evidence against a threshold effect rather than as a signal of misspecification.

3.4.4 Spatial Durbin Model

A Spatial Durbin Model (SDM) is estimated to characterise the spatial structure of the associations between capacity composition and energy deficit across state borders. The SDM specification is:

$$Y = lWY + Xb + WXq + m + e$$

where W is the row-standardised spatial weight matrix, l is the spatial autoregressive parameter capturing direct outcome-to-outcome dependence across neighbours, b captures the direct own-state associations of the two capacity variables, q captures indirect spillover associations of neighbouring states' capacity variables, m is the vector of individual state fixed effects, and e is the idiosyncratic error. The SDM nests both the spatial lag model (q = 0) and the spatial error model (l = 0, q = -lb) as special cases, making it the preferred general specification when the source of spatial dependence is not known a priori (Elhorst, 2014; LeSage and Pace, 2009).

Two spatial weight matrices are constructed. The primary matrix assigns a value of one to any pair of states belonging to the same regional grid zone, Northern, Western, Southern, Eastern, or North-Eastern, as recorded in the installed capacity source data, and zero otherwise. This definition reflects the institutional and physical architecture of inter-state power flows in India, where regional grid interconnections determine which states can dispatch surplus generation to or receive emergency power from one another during shortfall periods. The robustness matrix uses geographic contiguity, assigning a value of one to states sharing a land border and zero otherwise. Island territories without land borders are assigned to their nearest mainland state by established grid and logistics ties: Andaman and Nicobar Islands to Tamil Nadu and Lakshadweep to Kerala. Both matrices are row-standardised such that each row sums to unity. Moran's I tests are conducted on the primary FE model residuals, computed separately for each sample year under the grid-region weight matrix, to establish whether spatial autocorrelation is present in the outcome variable independent of the full spatial model.

LeSage-Pace Impacts Decomposition. The raw coefficients from SDM estimation do not represent the true marginal associations of the regressors because they ignore the spatial multiplier process induced by the spatial lag parameter l. Following LeSage and Pace (2009), the total, direct, and indirect effects of each regressor are computed from the impacts matrix S(W) = (I - lW)^{-1}. For a regressor k with own-state coefficient bk and spatially lagged coefficient qk:

$$Sk(W) = (I - lW)^{-1} (bk * I + qk * W)$$

The average direct effect is the mean diagonal element of Sk(W), capturing the own-state association net of spatial feedback. The average indirect effect is the mean row-sum minus the mean diagonal element, capturing the net spillover association onto all other states. The average total effect is the mean row-sum. Standard errors for the impact estimates are obtained by 300-draw Monte Carlo simulation from the asymptotic distribution of the model parameters, implemented via Cholesky decomposition of the parameter covariance matrix. Both the grid-region and contiguity weight matrix specifications are subjected to the full impacts decomposition; robustness of the indirect effect across both specifications is treated as the criterion for a reliable spatial finding.

3.4.5 Robustness Checks

Five checks assess the stability of the primary finding across alternative outcome measures, sample restrictions, estimators, and specifications.

Check 1: Alternative reliability measure. The primary two-way FE specification is re-estimated with outage rate in place of energy deficit as the dependent variable, on 130 observations after accounting for missingness in the outage data. This tests whether the coal-reliability association holds under an operationalisation of reliability that captures unit-level operational disruption rather than aggregate system-level shortfall.

Check 2: Coal consumption subsample. The primary specification is re-estimated on the subsample of 44 state-years across 23 states in 2020 and 2021 for which coal consumption data are available, with coal consumption included as an additional covariate. The restricted coverage of the coal consumption series limits this check to two years and 23 states; results are reported as inconclusive if no coefficient reaches conventional significance given the reduced statistical power.

Check 3: Panel quantile regression. A least-squares dummy variable quantile regression is estimated at the 25th, 50th, and 75th percentiles of the energy deficit distribution, with state and year fixed effects included as dummy regressors and bootstrap standard errors based on 200 replications. This tests whether the coal-reliability association is concentrated at specific points of the deficit distribution or holds across its range.

Check 4: Relative importance decomposition. A Shapley value decomposition via the LMG sequential averaging method partitions the primary model's within-R-squared between the two capacity variables, using data transformed by the two-way within estimator. This quantifies the relative informational contribution of each regressor within the model's explained variance.

Check 5: Mundlak-CRE coefficient stability. The CRE estimates of b1 and b2 are compared directly against the FE point estimates. A material change in sign or magnitude would indicate that the FE associations are sensitive to the treatment of individual heterogeneity; stability across the two specifications is treated as evidence of robustness.

3.5 Software

All data processing and estimation are conducted in R version 4.4. Panel data structures, fixed effects, and random effects estimation use the plm package (Croissant et al., 2025). Cluster-robust and Driscoll-Kraay standard errors use the sandwich and lmtest packages. Cross-sectional dependence tests use the plm package implementations of Pesaran (2004) and Frees (1995). Spatial weight matrices and the Spatial Durbin Model are estimated using the spdep and splm packages. LeSage-Pace impacts are computed via the manual matrix implementation described in Section 3.4.4 using covariance matrices extracted through the spatialreg package. Panel quantile regression uses the quantreg package. Relative importance decomposition uses the relaimpo package. The Goodman-Bacon decomposition is implemented manually following the 2x2 comparison framework in Goodman-Bacon (2021). Artificial intelligence-assisted text drafting was used in preparing this manuscript; the author retains full accountability for all content, analytical choices, and claims.

4. Empirical Results

4.1 Panel Overview and Descriptive Patterns

The estimation sample covers 35 Indian states and union territories over 2020-2023, a balanced panel of 140 state-year observations. Energy deficit as a share of annual requirement ranges from near-zero in chronically well-supplied states to 6.4 percent in years of acute shortfall, with substantial cross-state variation in each year. Renewable capacity share ranges from zero in states with no installed renewable capacity to over 80 percent in hydro-dominant northeastern states, with a sample mean of 30.4 percent. Coal capacity share ranges from zero to 81.3 percent, with a sample mean of 29.7 percent. The two shares are negatively correlated ($r = -0.34$) but the correlation is far from perfect, indicating that high-coal and high-renewable states are not simply mirror images of each other. States with persistently high coal shares Chhattisgarh, Jharkhand, Odisha, West Bengal are coal-bearing states with legacy thermal infrastructure. States with persistently high renewable shares Rajasthan, Gujarat, Tamil Nadu reflect early renewable procurement policy and favourable solar and wind resource endowments.

The Goodman-Bacon (2021) decomposition confirms that the cross-sectional and within-state temporal patterns run in opposite directions. The weighted average of 306 pairwise 2x2 difference-in-differences comparisons based on a median-split coal capacity proxy is +0.601: persistently high-coal states tend on average to show higher deficits than persistently low-coal states. The within-state FE coefficient is -0.102: within each state, years of higher coal share are associated with lower deficits. The TWFE estimator recovers the within-state temporal association; all results in this section carry that scope restriction. Of the 306 state-pair comparisons, 130 (42.5 percent) produce negative estimates, confirming substantial treatment effect heterogeneity across state pairs.

4.2 Cross-Sectional Dependence Diagnostics

Prior to estimation of the primary model, two formal tests assess cross-sectional dependence in the panel residuals. The Pesaran (2004) CD test rejects the null of no cross-sectional dependence ($z = 4.493$, $p < 0.001$). The Frees (1995) test returns a positive average pairwise residual correlation (rho = 0.092). The mean absolute pairwise correlation computed directly across all 595 state pairs is 0.489. These results confirm that approximately half the variation in model residuals is contemporaneously shared across states, consistent with common shocks affecting the Indian electricity grid simultaneously across multiple states. State-level clustered standard errors, which account for within-state serial dependence but not contemporaneous cross-state dependence, are therefore insufficient for inference. Driscoll-Kraay (1998) standard errors, which are robust to all three forms of dependence, are used as the primary basis for inference in all models that follow. The detected cross-sectional dependence provides pre-estimation structural evidence that a spatial modelling framework is warranted.

4.3 Primary Panel Estimates

The table 1 reports the two-way fixed effects model under both standard error corrections. Coal capacity share carries a negative and significant coefficient under both standard error corrections: $p = 0.035$ under clustering and $p = 0.002$ under the preferred Driscoll-Kraay correction. The strengthening of significance from clustered to Driscoll-Kraay standard errors reflects the structure of the cross-sectional dependence: contemporaneous cross-state error correlation was inflating clustered standard errors. Within-state increases in coal capacity share are associated with lower state-level energy deficits, conditional on state and year effects. Renewable capacity share shows no detectable within-state association with energy deficit under either correction.

Table 1 Two-Way Fixed Effects Model, Dependent Variable: Energy Deficit Percentage

Variable	Coef.	Cluster SE	p	DK SE	p
Renewable capacity share	-0.00099	0.01048	0.925	0.00100	0.328
Coal capacity share	-0.10150	0.04756	0.035**	0.03156	0.002***

*Note. ** $p < 0.05$; *** $p < 0.01$. Balanced panel: $n = 35$, $T = 4$, $N = 140$. State and year two-way fixed effects. Cluster SE: clustered at state level. DK SE: Driscoll-Kraay (1998) standard errors, robust to cross-sectional dependence, heteroskedasticity, and serial correlation. Within R-squared: 0.015. F-test (two-way FE versus pooled OLS): $F = 3.31$, $p < 0.001$.*

The coal coefficient magnitude, 0.10 percentage points of deficit reduction per percentage point increase in coal capacity share, is modest and reflects capacity composition as a partial signal within a model that explains 1.5 percent of total outcome variation. The dominant determinants of state-level reliability, including discom financial health, AT&C losses, and demand-side variability, lie outside the specification (Verma et al., 2020; Power Finance Corporation, 2024).

Three specification tests support the fixed effects choice. The F-test rejects pooled OLS ($F = 3.31$, $p < 0.001$). The Hausman test fails to reject consistency of random effects (chi-squared = 2.16, df = 2, $p = 0.340$). The Mundlak-Chamberlain CRE Wald test fails to reject the joint null that the Mundlak correction terms equal zero (chi-squared = 3.797, df = 2, $p = 0.150$), confirming formally that the individual state effects are not systematically correlated with the regressors. These three results indicate that fixed and random effects are statistically equivalent in this panel; the FE specification is retained on unconditional consistency grounds.

4.4 Mundlak-Chamberlain Correlated Random Effects

The table 2 reports the CRE model alongside the FE point estimates to assess coefficient stability across specifications. The coal capacity coefficient attenuates from -0.102 in the FE model to -0.071 in the CRE model, a reduction of approximately 30 percent in absolute magnitude while retaining the same sign. This attenuation is consistent with mild positive correlation between the state-level coal capacity mean and the individual effect, but the Wald test confirms this correlation is not statistically distinguishable from zero at conventional levels.

Table 2 Coefficient Stability: Two-Way FE versus Mundlak-Chamberlain CRE

Variable	FE Coef.	FE p (DK)	CRE Coef.	CRE p (Cluster)
Renewable capacity share	-0.00099	0.328	-0.00077	0.944
Coal capacity share	-0.10150	0.002***	-0.07142	0.162
Mean renewable share (delta1)	—	—	0.02560	0.059†
Mean coal share (delta2)	—	—	0.07869	0.144
Wald test: delta1 = delta2 = 0	—	—	chi2 = 3.797	0.150

*Note. *** $p < 0.01$; † $p < 0.10$. FE p-values based on Driscoll-Kraay standard errors. CRE p-values based on cluster-robust standard errors at the state level. CRE R-squared: 0.018.*

The mean renewable capacity share (delta1) is individually marginal at $p = 0.059$, suggesting states with persistently higher renewable shares may have some unobserved structural characteristic that independently affects deficits; the joint test does not support this interpretation at the 10 percent level, and the renewable coefficient itself remains near zero and insignificant in both specifications. The coal finding is directionally stable; the renewable null is stable.

4.5 Threshold Nonlinearity Test

The Hansen (1999) grid search over the 5th to 95th percentile of the renewable generation share distribution converges to an estimated threshold of gamma = 0.011. This value sits at the lower boundary of the search grid rather than at an interior break point. The bootstrap test for threshold significance, based on 300 replications, returns $p = 0.163$, failing to reject the null of no threshold effect. The boundary-degenerate estimate and null bootstrap result are jointly interpreted as the

absence of a detectable interior regime shift in the renewable capacity-reliability association within the observed range of renewable generation share. The linear two-way FE model of Section 4.3 is the preferred specification.

4.6 Spatial Durbin Model

4.6.1 Pre-Estimation Spatial Diagnostics

Moran's I tests on the primary FE residuals, computed separately for each sample year under the grid-region weight matrix, find no significant spatial autocorrelation in any year: 2020 (I = -0.120, p = 0.879), 2021 (I = -0.032, p = 0.512), 2022 (I = -0.036, p = 0.528), 2023 (I = +0.016, p = 0.258). The outcome variable, energy deficit, does not cluster spatially across state borders. This result rules out direct outcome contagion as the source of the cross-sectional dependence detected in Section 4.2, pointing instead to shared common shocks and potential covariate-channel spillovers as the operative mechanisms.

4.6.2 SDM Raw Coefficients

Table 3 reports the SDM estimates under both weight matrix specifications. Raw SDM coefficients are not marginal effects; the correct direct, indirect, and total effects are reported in Table 4.

Table 3 Spatial Durbin Model: Raw Coefficients, Dependent Variable: Energy Deficit Percentage

Variable	Grid-Region Coef.	Grid-Region p	Contiguity Coef.	Contiguity p
Spatial lag (lambda)	-0.250	0.153	0.009	0.935
Renewable capacity share	-0.00067	0.883	-0.00114	0.825
Coal capacity share	-0.10311	0.112	-0.11477	0.085†
W-renewable capacity share	0.01132	0.203	0.01421	0.165
W-coal capacity share	0.30826	0.012**	0.17802	0.034**

*Note. ** $p < 0.05$; † $p < 0.10$. Individual state fixed effects. N = 140. W-prefix denotes spatially lagged variable. Raw SDM coefficients conflate own-state associations with spatial feedback and are not marginal effects; see Table 4 for the LeSage-Pace impacts decomposition.*

The spatial autoregressive parameter lambda is statistically indistinguishable from zero under both weight specifications (p = 0.153 and p = 0.935), confirming the Moran's I diagnostic: deficits do not transmit directly across state borders. The spatially lagged coal capacity share is significant and positive under both weight matrices. The spatially lagged renewable capacity share is not significant under either specification.

4.6.3 LeSage-Pace Impacts Decomposition

Table 4 reports the average direct, indirect, and total effects recovered from the LeSage-Pace (2009) impacts matrix, with standard errors from 300-draw Monte Carlo simulation.

Table 4 Spatial Durbin Model: LeSage-Pace Direct, Indirect, and Total Impacts

Variable / Effect	Grid-Region Direct	Grid-Region Indirect	Grid-Region Total	Contiguity Direct	Contiguity Indirect	Contiguity Total
Renewable capacity share	-0.00107	0.00959	0.00852	-0.00112	0.01429	0.01318
p-value	0.822	0.203	0.254	0.837	0.167	0.164
Coal capacity share	-0.11479	0.27893	0.16414	-0.11443	0.17824	0.06381
p-value	0.089†	0.010**	0.099†	0.056†	0.024**	0.454

*Note. ** p < 0.05; † p < 0.10. Standard errors from 300-draw Monte Carlo simulation via Cholesky decomposition of the parameter covariance matrix. Direct effect: mean diagonal of S(W) = (I - lambdaW)^{-1}(beta*I + theta*W). Indirect effect: mean row-sum minus mean diagonal. Total: mean row-sum.*

Three findings emerge from the impacts decomposition.

First, the direct effect of coal capacity share on own-state energy deficit is negative under both weight matrices (-0.115 under grid-region, p = 0.089; -0.114 under contiguity, p = 0.056). Within-state increases in coal capacity share are marginally associated with lower own-state energy deficits, consistent with the primary FE finding.

Second, the indirect effect of coal capacity share is positive and significant under both weight matrices (0.279, p = 0.010 under grid-region; 0.178, p = 0.024 under contiguity). A state's energy deficit rises with the coal capacity share of its spatial neighbours, independent of its own coal capacity share. This indirect effect is robust across two structurally different definitions of spatial proximity and constitutes the paper's principal spatial finding.

Third, neither the direct nor indirect effects of renewable capacity share reach significance under any impact category or weight matrix specification. No detectable spatial association between renewable capacity composition and energy deficit exists in this panel.

The positive sign of the coal indirect effect is counter-intuitive relative to a simple inter-state dispatch hypothesis. Two mechanisms are consistent with the observed positive sign. States with high coal capacity shares may consume the majority of their own thermal output during peak demand periods, leaving little surplus available for export to neighbouring states. Alternatively, the spatial lag may partly proxy for unobserved demand-side clustering within regional grid zones: states in the same zone tend to face similar industrial demand structures, and neighbouring states' coal capacity may reflect shared demand-side pressure rather than supply-side interconnection. The identification strategy cannot separate these mechanisms, and both should be considered in interpreting this result.

4.7 Robustness Checks

Table 5 reports coefficient estimates across all five robustness checks.

Check 1 (alternative reliability measure). Neither coefficient is significant when outage rate replaces energy deficit as the dependent variable. The coal-reliability association found for aggregate energy deficit does not replicate for unit-level outage rates. Energy deficit reflects aggregate supply adequacy governed by capacity composition and dispatch structure; outage rate reflects unit-level operational disruption driven by maintenance, fuel supply chain, and plant-level factors that capacity composition does not directly govern. The finding is scoped accordingly to aggregate supply adequacy.

Table 5 Robustness Checks: Summary of Coefficient Estimates

Check	**N**	**Variable**	**Coef.**	**p**
1: Outage rate model	130	Renewable capacity share	-0.033	0.151
		Coal capacity share	0.781	0.312
2: Coal subsample	44	Renewable capacity share	0.095	0.256
		Coal capacity share	-0.070	0.418
		Coal consumption	0.019	0.229
3: Quantile Q25	140	Renewable capacity share	-0.000003	0.999
		Coal capacity share	-0.033	0.356
3: Quantile Q50	140	Renewable capacity share	-0.000169	0.939
		Coal capacity share	-0.049	0.112
3: Quantile Q75	140	Renewable capacity share	0.000227	0.917
		Coal capacity share	-0.039	0.439
4: LMG decomposition	—	Coal share of within-R2	98.1%	—
		Renewable share of within-R2	1.9%	—
		Total within-R2	1.53%	—
5: Mundlak-CRE stability	140	Renewable capacity share	-0.00077	0.944
		Coal capacity share	-0.07142	0.162

Note. Check 1: two-way FE, cluster-robust SE, outage rate as dependent variable. Check 2: individual FE, cluster-robust SE, restricted to 2020-2021 state-years with non-missing coal consumption. Check 3: LSDV quantile regression with state and year dummies, bootstrap SE (200 replications); nonunique solution warnings returned at all three quantiles. Check 4: Shapley value decomposition via LMG method, two-way within-transformed data. Check 5: Mundlak-Chamberlain CRE, cluster-robust SE; see Table 2 for full specification.

Check 2 (coal consumption subsample). No coefficient reaches significance in the 44-observation subsample. The result is inconclusive: restricting to 2020-2021 removes two of four sample years, reduces the panel to 23 states, and introduces three collinear energy-side regressors simultaneously, reducing statistical power below the level needed to detect effects of the magnitude found in the primary model.

Check 3 (panel quantile regression). The coal capacity share coefficient is negative at all three quantiles (Q25 = -0.033, Q50 = -0.049, Q75 = -0.039) but statistically insignificant throughout. The direction is consistent with the primary FE finding across the deficit distribution. Nonunique solution warnings at all three quantiles, reflecting the ratio of free parameters to observations under full state and year dummy saturation, indicate that these estimates should be treated as directional rather than confirmatory.

Check 4 (relative importance). The LMG Shapley decomposition confirms that coal capacity share accounts for 98.1 percent of the variance explained by the two-variable primary model. The total within-R-squared is 1.53 percent. Coal dominates the model's explained variance entirely within a model that itself accounts for a small fraction of total outcome variation.

Check 5 (Mundlak-CRE stability). The coal coefficient attenuates from -0.102 to -0.071 across the FE and CRE specifications, a sign-preserving reduction of approximately 30 percent. The renewable coefficient is stable at zero. Directional stability across both specifications, combined with the non-significant Wald test on the Mundlak correction terms, confirms that the FE estimates are not sensitive to the treatment of individual heterogeneity.

4.8 Synthesis of Results

Table 6 summarises the sign and significance of the coal capacity share and renewable capacity share associations across all estimated models.

Table 6 Sign and Significance of Capacity Share Associations Across All Models

Model	N	Coal: sign	Coal: sig.	Renewable: sign	Renewable: sig.
Two-way FE (Cluster SE)	140	Negative	p=0.035**	Negative	p=0.925
Two-way FE (DK SE)	140	Negative	p=0.002***	Negative	p=0.328
Mundlak-CRE	140	Negative	p=0.162	Negative	p=0.944
SDM direct (Grid-Region W)	140	Negative	p=0.089†	Negative	p=0.822
SDM direct (Contiguity W)	140	Negative	p=0.056†	Negative	p=0.837
SDM indirect (Grid-Region W)	140	Positive	p=0.010**	Positive	p=0.203
SDM indirect (Contiguity W)	140	Positive	p=0.024**	Positive	p=0.167
Quantile Q25	140	Negative	p=0.356	Negative	p=0.999
Quantile Q50	140	Negative	p=0.112	Negative	p=0.939
Quantile Q75	140	Negative	p=0.439	Positive	p=0.917
Outage rate (Check 1)	130	Positive	p=0.312	Negative	p=0.151
Coal subsample (Check 2)	44	Negative	p=0.418	Positive	p=0.256

*Note. ** $p < 0.05$; *** $p < 0.01$; † $p < 0.10$. SDM indirect effects from Table 4. Mundlak-CRE from Table 2. All associations are correlational; the identification strategy does not support causal claims.*

Four conclusions follow from Table 6. First, the coal capacity share direct association with own-state energy deficit is negative in ten of twelve model-sample combinations and significant under

the preferred Driscoll-Kraay correction, the SDM direct effect, and the primary FE model. Second, the coal indirect effect is positive and significant under both spatial weight specifications, the paper's most robust finding in terms of statistical precision and cross-specification consistency. Third, renewable capacity share shows no detectable association with energy deficit in any direction, at any quantile, or under any spatial specification across all twelve estimation environments.

Fourth, the coal direct effect does not replicate for the outage-based reliability measure, confirming that the finding is scoped to aggregate supply adequacy rather than to grid reliability defined more broadly. All associations in this paper are correlational; the identification strategy does not support causal claims.

5. Discussion

5.1 The Coal-Reliability Association: Scope and Interpretation

The primary finding is that within-state increases in coal capacity share are associated with lower state-level energy deficits, conditional on state and year effects. This association carries four scope conditions that together determine how it may be used.

First, the association is correlational. The two-way FE estimator recovers a within-state temporal association, but coal capacity share is not randomly assigned. States with persistently higher coal shares are coal-bearing states with historically lower power procurement costs, stronger discom balance sheets, and stronger industrial demand bases, each of which independently predicts lower deficits. The Mundlak-Chamberlain CRE Wald test (chi-squared = 3.797, p = 0.150) confirms that these state-level characteristics are not statistically correlated with the within-state variation in coal capacity share during the sample period, supporting the FE specification but not converting it into a causal identification. The Goodman-Bacon decomposition adds a further dimension: the cross-sectional pattern runs in the opposite direction from the within-state temporal pattern. States that are persistently coal-heavy tend on average to show higher deficits than persistently low-coal states, a pattern consistent with coal-heavy states clustering in lower-income regions with chronically stressed discoms (Verma et al., 2020). The within-state FE coefficient identifies a different relationship, the year-on-year co-movement of coal capacity share and energy deficit within each state, which runs negative. These two patterns are not contradictory; they operate at different levels of the data structure and should be read as distinct empirical regularities rather than as conflicting evidence about the same underlying relationship.

Second, the association applies specifically to aggregate energy deficit and does not replicate when reliability is measured through unit-level outage rates. Energy deficit reflects supply adequacy at the system level, which coal's dispatchable baseload characteristics directly support. Outage rates reflect operational reliability at the generating-unit level, which maintenance schedules, fuel supply

chain continuity, and plant-level management govern rather than aggregate capacity composition. The coal finding is scoped accordingly.

Third, the LeSage-Pace impacts decomposition reveals a spatial dimension not visible from the primary FE coefficient alone. The direct effect of own-state coal capacity on own-state deficit is negative and marginally significant under both weight matrix specifications. The indirect effect of neighbouring states' coal capacity share on own-state deficit is positive and significant at $p = 0.010$ under the grid-region weight matrix and $p = 0.024$ under geographic contiguity. A state's deficit rises as the coal capacity share of its regional grid neighbours rises. Two mechanisms are consistent with this positive indirect effect. Coal-heavy neighbours may consume the majority of their own thermal output during peak demand periods, leaving little surplus dispatchable power available for inter-state transfer to deficit states. Alternatively, states within the same regional grid zone may face similar industrial demand structures, and neighbouring coal capacity may partly proxy for shared demand-side pressure rather than reflecting supply-side grid interdependence. The identification strategy cannot distinguish between these mechanisms.

Fourth, the Mundlak-CRE attenuation, coal coefficient moving from -0.102 to -0.071 across the two specifications, indicates that part of the FE coefficient reflects within-state variation mildly correlated with persistent state-level characteristics. The sign is preserved, but the magnitude should be treated as an upper bound on the within-state association rather than as a precisely calibrated effect size.

5.2 The Renewable Capacity Null: Structural Interpretation

Renewable capacity share shows no detectable association with energy deficit across twelve estimation environments. The consistency of the null across this range argues against a statistical artifact attributable to any single estimation choice and points toward a structural feature of the Indian power system during the sample period.

The national capacity-generation gap provides the structural explanation. As of late 2025, renewable and coal capacity shares approach parity in India's installed base, yet renewable generation accounts for only approximately 29 percent of total electricity produced (Manojkumar et al., 2026). This divergence reflects the intermittency of solar and wind output, their position behind coal in the dispatch order under binding long-term power purchase agreements, and the absence of grid-scale storage and flexible demand infrastructure that would allow variable renewable generation to be absorbed during peak load hours. A state can expand renewable installed capacity substantially without gaining a proportional increase in the dispatchable supply that determines whether aggregate demand is met during shortfall events. Li et al. (2025) document an equivalent spatiotemporal mismatch for Chinese provinces, driven by the same combination of grid constraints, dispatch-order disadvantages, and storage limitations. Lu et al. (2022) demonstrate that

addressing this mismatch requires expansion of inter-regional balancing areas rather than plant-level retrofits.

The Hansen threshold test adds a specific finding within this null. The grid search converges to a boundary-degenerate threshold estimate at the lower edge of the search range, and the bootstrap p-value is 0.163. The data contain no interior break point in the renewable capacity-reliability relationship within the observed range of renewable generation share. Whether a threshold effect exists beyond the penetration levels reached in this sample is a question the current four-year panel cannot address.

5.3 Positioning Within the SDG 7 Framework

The results speak to three of SDG 7's sub-targets with varying degrees of directness.

On indicator 7.1.1, access to electricity, India has functionally achieved universal household connection under the Saubhagya program, placing it ahead of the global electrification trajectory documented in successive editions of Tracking SDG 7 (IEA et al., 2024; IEA et al., 2025). The panel results in this paper address the quality dimension of access that indicator 7.1.1 does not track: whether supply meets demand once a household or firm is connected to the grid. Measurable energy deficits persist across 35 states in the 2020-2023 sample. The persistence is uneven across states in ways that national-level indicators cannot surface, and that subnational panel analysis is specifically designed to document.

On indicator 7.2.1, the share of renewable energy in total final energy consumption, India has achieved rapid progress on an installed capacity basis, reaching fifty percent non-fossil installed capacity five years ahead of its NDC target. The null renewable reliability finding in this paper is consistent with the evidence that renewable generation share has not kept pace with renewable capacity share nationally (Manojkumar et al., 2026). The Tracking SDG 7 report notes that capacity-based and generation-based indicators can diverge significantly (IEA et al., 2024); the state-level panel results here corroborate that divergence in a reliability context. Monitoring indicator 7.2.1 through installed capacity share alone may overstate the operational contribution of renewable energy to supply adequacy and to the reliability dimension of SDG 7.1.

On target 7.b, the planned infrastructure expansion of over 190,000 circuit kilometres of transmission through fiscal year 2032 is central to enabling renewable dispatch and reducing curtailment (Manojkumar et al., 2026). This paper cannot test whether transmission adequacy moderates the capacity-reliability relationships because state-level transmission data are not available at the required spatial resolution. The coal spatial spillover finding points indirectly to the same bottleneck: inter-state transmission currently does not efficiently route surplus coal generation to deficit states within regional grid zones, consistent with the dispatch and congestion constraints identified in the curtailment literature for China and Germany (Lu et al., 2022; Shu et al., 2020).

These findings also connect to the IPCC AR6 Working Group III assessment, which identifies inter-state transmission expansion as a necessary condition for renewable integration at sub-national scale (IPCC, 2022). The 2020-2023 sample period corresponds to a stage of India's energy transition in which renewable capacity has expanded rapidly but the transmission and storage infrastructure identified in AR6 scenarios as necessary for reliable dispatch has not yet materialised at scale. The null renewable reliability finding and the coal covariate spillover are consistent with being at that stage of the transition pathway, not with renewable energy being inherently unreliability-increasing.

5.4 Methodological Contributions and Extensions

The paper makes three methodological contributions beyond its substantive findings.

The grid-region spatial weight matrix, constructed from India's five official regional grid zones rather than from geographic adjacency, is a theoretically grounded definition of spatial proximity that differs from the standard distance-based or contiguity-based matrices used in most subnational energy studies. The robustness of the coal indirect effect across both weight matrix specifications strengthens confidence that the finding reflects genuine cross-state interdependence rather than an artefact of a particular spatial structure choice.

The LeSage-Pace impacts decomposition separates the direct and indirect effects of each regressor through the spatial multiplier $S(W) = (I - \rho W)^{-1}$, revealing a finding not visible from raw SDM coefficients. The raw W-coal coefficient in Table 3 is 0.308 under grid-region weighting; the indirect impact from Table 4 is 0.279. The difference reflects the correction for spatial feedback through S(W). Reporting raw SDM coefficients as marginal effects, a practice common in applied spatial panel papers, would have slightly overstated the indirect association in this case.The sequential application of the Goodman-Bacon decomposition, the Mundlak-CRE specification, the Pesaran CD test, and the Driscoll-Kraay standard error correction provides a diagnostic sequence more comprehensive than the standard Hausman test approach used in most panel studies of energy outcomes. Each diagnostic has substantive implications for how the primary coefficient is interpreted; reporting them is methodologically necessary rather than merely thorough.

The natural methodological extension is a panel vector autoregression (PVAR) applied to a longer panel once post-2023 data become available (Kazemzadeh et al., 2023). With T = 4, the static specification cannot recover the dynamic feedback between capacity investment decisions and subsequent reliability outcomes. The PVAR framework would allow the bidirectional relationship to be estimated, testing whether renewable capacity additions reduce deficits with a lag once sufficient within-state variation has accumulated across a longer time horizon.

5.5 Limitations

The seven limitations bound the interpretation of these results, stated here as a consolidated account.

The identification is correlational throughout. No instrument, natural experiment, or treatment assignment mechanism is available to support causal claims. The Goodman-Bacon decomposition confirms substantial treatment effect heterogeneity across state pairs (42.5 percent of comparisons produce negative estimates), and the Callaway-Sant'Anna (2021) heterogeneity-robust estimator cannot be applied given the continuous treatment variable and T = 4.

The panel covers four years. Cross-sectional dependence at this scale prevents two-way random effects estimation. The short time dimension limits the power of both the Hansen threshold test and the spatial estimators.

The coal direct effect does not replicate for the outage-based reliability measure. The finding is scoped to aggregate supply adequacy and cannot be generalised to grid reliability defined through unit-level operational continuity.

The coal consumption subsample check (N = 44, two years, 23 states) is inconclusive due to sample size. Coal consumption data terminate in 2021, preventing a current-period sensitivity analysis using this variable.

The quantile regression estimates carry nonunique solution warnings at all three quantiles, reflecting the ratio of free parameters to observations under full state and year dummy saturation. They are treated as directional evidence rather than confirmatory results.

The model's within-R-squared is 1.53 percent. The dominant determinants of state-level energy deficits, including discom financial health, AT&C losses, and demand-side variability, lie outside the specification because state-year panel data on these variables are not available at the coverage and frequency of the primary outcome data.

The positive sign of the coal indirect effect is consistent with two mechanisms that the identification strategy cannot separate. The finding is reported as a descriptive spatial association, not as evidence of a specific inter-state dispatch or demand-clustering process.

6. Policy Implications

6.1 State-Level Power Planning

The null renewable capacity finding holds across twelve estimation environments and two reliability measures. It does not imply that renewable capacity investment is ineffective. It implies that capacity additions without complementary investment in grid-scale storage, demand flexibility, and transmission infrastructure have not translated into measurable aggregate deficit reductions at the state level during 2020-2023. The capacity-generation gap visible at the national level is present at the state level and is not resolved by examining the data at finer spatial resolution.

States face a sequencing decision rather than a trade-off. Renewable capacity expansion and the infrastructure investments needed for reliable dispatch are complements, not substitutes. States that have expanded renewable capacity rapidly without parallel storage procurement and time-of-day

tariff infrastructure face a reliability dividend that is deferred rather than delivered. The evidence from China's provincial experience indicates that expanding balancing areas and inter-regional transmission, not retrofitting thermal plants for greater flexibility, is the operationally effective path to reducing curtailment and realising reliability gains from renewable deployment (Lu et al., 2022). India's planned transmission buildout of over 190,000 circuit kilometres through fiscal year 2032 is the relevant instrument; the question this paper cannot yet answer is whether states receiving disproportionate shares of that investment show faster convergence between renewable capacity share and renewable generation share.

The coal indirect effect carries a specific planning implication. A state's energy deficit rises with the coal capacity share of its regional grid neighbours, robust to grid-region and geographic contiguity definitions of spatial proximity. The inter-state reliability buffer that operates through regional grid interconnection currently functions through coal-based dispatch. Renewable capacity in neighbouring states does not yet replicate this function. Power planners in deficit-prone states who rely on regional interconnection as a reliability backstop should recognise that this backstop is currently a coal-based mechanism, and that its availability during peak demand periods will depend on the surplus dispatch capacity of neighbours.

6.2 SDG 7 Monitoring Frameworks

The installed capacity share is the primary metric used to track progress on SDG 7.2.1, and the Tracking SDG 7 reports acknowledge that capacity-based and generation-based indicators can diverge (IEA et al., 2024). The state-level panel results in this paper provide empirical evidence of that divergence's reliability consequence: renewable capacity at current penetration levels carries no detectable association with the supply adequacy dimension of SDG 7.1 at the state level. A monitoring framework that reports only capacity milestones cannot surface this gap.

Two specific refinements follow from this finding. State-level generation share, computed as renewable generation divided by total generation rather than as renewable capacity divided by total capacity, should be tracked as a companion indicator to SDG 7.2.1. It measures whether installed renewable capacity is being converted into dispatched electricity rather than remaining idle due to curtailment or scheduling constraints. State-level energy deficit, the variable constructed and analysed in this paper from publicly available Central Electricity Authority data, is already reported monthly and could be integrated into India's subnational SDG tracking framework. The Tracking SDG 7 reports document global and national progress; a subnational layer would allow the between-state dispersion that this paper documents to be monitored systematically.

6.3 Transmission, Storage, and Infrastructure Planning

The inability to identify a state-level transmission effect in this paper is a data constraint rather than a substantive finding: the source data record transmission line projects by route pair rather than by

state, preventing attribution of transmission investment to specific states. State-level transmission and storage capacity data are the most important missing variables for extending this analysis, and their systematic collection as part of India's energy sector administrative data infrastructure would directly enable the next round of research questions this paper identifies. The Revamped Distribution Sector Scheme, the National Electricity Plan, and the state-level renewable purchase obligation framework all generate implementation data that, if compiled at the state-year level, would allow direct testing of whether transmission and storage investment moderates the capacity-reliability associations identified here.

The coal spatial spillover finding points to a specific infrastructure question this paper cannot answer but that future state-level data would address. If the positive indirect effect of neighbouring coal capacity reflects limited inter-state transfer of surplus coal generation to deficit states, then expanding high-voltage direct current inter-state transmission links within each regional grid zone would be expected to turn this effect negative, as surplus coal generation from coal-heavy states becomes more accessible to deficit-prone neighbours. If the positive indirect effect reflects demand-side clustering within grid zones rather than supply-side grid constraints, transmission investment would not change the sign of the indirect effect. Distinguishing between these interpretations requires state-level power flow data not currently available in the public domain.

7. Conclusion

This paper examines the correlational association between electricity capacity composition and grid reliability across 35 Indian states from 2020 to 2023, deploying a two-way fixed effects model with Driscoll-Kraay standard errors, a Mundlak-Chamberlain correlated random effects specification, a Goodman-Bacon decomposition, a Pesaran cross-sectional dependence test, a Hansen panel threshold test, and a Spatial Durbin Model with LeSage-Pace direct, indirect, and total impacts decomposition under two spatial weight matrices.

The principal empirical findings are four. Coal capacity share is negatively associated with own-state energy deficit within states over time, an association that strengthens from $p = 0.035$ under state-clustered standard errors to $p = 0.002$ under Driscoll-Kraay standard errors that correct for the strong cross-sectional dependence documented in this panel (mean absolute pairwise residual correlation = 0.489). Renewable capacity share shows no detectable association with energy deficit in any of twelve estimation environments spanning five estimators, two reliability measures, and two spatial weight matrix specifications. The Hansen threshold test finds no interior break point in the renewable capacity-reliability relationship within the observed range of renewable generation share. The LeSage-Pace impacts decomposition reveals that coal capacity generates a positive and significant indirect effect on own-state deficit through neighbouring states' capacity composition, robust to both grid-region and geographic contiguity definitions of spatial proximity (indirect

impact = 0.279, p = 0.010 and 0.178, p = 0.024 respectively), while Moran's I tests confirm that the outcome itself carries no direct spatial autocorrelation across state borders.

Five qualifications are integral to these findings. The identification is correlational throughout; the Goodman-Bacon decomposition confirms that the TWFE coefficient captures a within-state temporal association that runs opposite to the cross-sectional pattern. The coal direct effect does not replicate for unit-level outage rates, scoping the finding to aggregate supply adequacy. The model explains 1.53 percent of total outcome variation, confirming that capacity composition is a partial signal within a reliability outcome dominated by discom financial health, AT&C losses, and demand-side variability. The coal indirect effect's positive sign is consistent with two mechanisms, limited surplus dispatch from coal-heavy neighbours versus demand-side clustering within grid zones, that the identification strategy cannot separate. The four-year panel window cannot test whether the null renewable relationship emerges at penetration levels beyond those observed.

Positioned against Sustainable Development Goal 7, the results describe the Indian energy transition at a specific stage: rapid expansion of installed renewable capacity without yet closing the gap between capacity buildout and reliable dispatch. India has functionally achieved the connection dimension of indicator 7.1.1 and has made rapid measurable progress on the capacity dimension of indicator 7.2.1. The supply reliability dimension of 7.1, whether energy reaches connected consumers when demanded, remains unevenly distributed across states in ways that national-level composite indicators do not surface. Progress on indicator 7.2.1, tracked through capacity share, overstates the operational contribution of renewable energy to supply adequacy at the state level during this sample period.

Future research in three directions would extend this contribution. A panel vector autoregression applied once the panel extends to seven or eight years would recover the dynamic feedback between capacity investment and reliability outcomes that the static specification cannot identify. Incorporation of state-year data on discom financial health, AT&C losses, and storage and transmission investment would address the specification's dominant limitation and test whether the capacity-reliability associations identified here are moderated by grid infrastructure quality. Extension of the spatial framework to include explicit inter-state power flow data from load despatch centre records would allow the direct identification of whether the coal indirect effect reflects supply-side dispatch constraints or demand-side clustering, resolving the interpretive ambiguity that the current specification leaves open.

References

Azaliah, N., Kurniawan, R., Hartono, D., & Widyastaman, P. (2023). The convergence of energy intensity in developing countries: A spatial econometric analysis with Indonesia's provincial panel data. Environment, Development and Sustainability. Advance online publication. https://doi.org/10.1007/s10668-023-03227-8

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200-230. https://doi.org/10.1016/j.jeconom.2020.12.001

Chamberlain, G. (1982). Multivariate regression models for panel data. Journal of Econometrics, 18(1), 5-46. https://doi.org/10.1016/0304-4076(82)90094-X

Croissant, Y., Millo, G., & Tappe, K. (2025). plm: Linear Models for Panel Data. R package version 2.6-4. https://CRAN.R-project.org/package=plm

Davidson, M. R., Zhang, D., Xiong, W., Zhang, X., & Karplus, V. J. (2016). Modelling the potential for wind energy integration on China's coal-heavy electricity grid. Nature Energy, 1(7), 16086. https://doi.org/10.1038/nenergy.2016.86

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. Review of Economics and Statistics, 80(4), 549-560. https://doi.org/10.1162/003465398557825

Elhorst, J. P. (2014). Spatial econometrics: From cross-sectional data to spatial panels. Springer. https://doi.org/10.1007/978-3-642-40340-8

Ember. (2025). India electricity data. Ember Energy. https://ember-energy.org/countries-and-regions/india/

Frees, E. W. (1995). Assessing cross-sectional correlation in panel data. Journal of Econometrics, 69(2), 393-414. https://doi.org/10.1016/0304-4076(94)01658-M

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254-277. https://doi.org/10.1016/j.jeconom.2021.03.014

Gulagi, A., Ram, M., Bogdanov, D., Sarin, S., Mensah, T. N. O., & Breyer, C. (2022). The role of renewables for rapid transitioning of the power sector across states in India. Nature Communications, 13, 5499. https://doi.org/10.1038/s41467-022-33048-8

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. Journal of Econometrics, 93(2), 345-368. https://doi.org/10.1016/S0304-4076(99)00025-1

Hausman, J. A. (1978). Specification tests in econometrics. Econometrica, 46(6), 1251-1271. https://doi.org/10.2307/1913827

IEA. (2024). Maintaining a stable electricity grid in the energy transition. International Energy Agency. https://www.iea.org/reports/maintaining-a-stable-electricity-grid-in-the-energy-transition

IEA, IRENA, UNSD, World Bank, & WHO. (2024). Tracking SDG 7: The energy progress report 2024. International Energy Agency. https://www.iea.org/reports/tracking-sdg7-the-energy-progress-report-2024

IEA, IRENA, UNSD, World Bank, & WHO. (2025). Tracking SDG 7: The energy progress report 2025. International Renewable Energy Agency. https://www.irena.org/Publications/2025/Jun/Tracking-SDG-7-The-Energy-Progress-Report-2025

IPCC. (2022). Climate change 2022: Mitigation of climate change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (P. R. Shukla et al., Eds.). Cambridge University Press. https://doi.org/10.1017/9781009157926

Kazemzadeh, E., Fuinhas, J. A., Koengkan, M., & Shadmehri, M. T. A. (2023). Relationship between the share of renewable electricity consumption, economic complexity, financial development, and oil prices: A two-step club convergence and PVAR model approach. International Economics, 173, 260-275. https://doi.org/10.1016/j.inteco.2022.12.003

LeSage, J. P., & Pace, R. K. (2009). Introduction to spatial econometrics. CRC Press. https://doi.org/10.1201/9781420064254

Li, Y., Chen, Y., Zhou, Y., & Wang, Z. (2025). Spatiotemporal dynamics and factors of renewable energy mismatch in China. Renewable and Sustainable Energy Reviews. https://doi.org/10.1016/j.rser.2025.115259

Lu, X., Yao, T., Fang, K., Pan, G., & McElroy, M. B. (2022). Large balancing areas and dispersed renewable investment enhance grid flexibility in a renewable-dominant power system in China. Proceedings of the National Academy of Sciences, 119(4), e2116632119. https://doi.org/10.1073/pnas.2116632119

Manojkumar, N., Mathew, M. M., & Myllyvirta, L. (2026). India power sector review 2025: Record clean energy deployment drives historic decline in coal generation. Centre for Research on Energy and Clean Air.

Ministry of Power, Government of India. (2024). AT&C losses: Performance data on distribution utilities. Rajya Sabha Unstarred Question No. 149, 25 November 2024. https://powermin.gov.in

Ministry of Power, Government of India. (2025). India's energy overview: Yearly highlights of 2024-25. Vasudha Foundation. https://www.vasudha-foundation.org/wp-content/uploads/Indias-Energy-Overview-year-end-report-of-FY-2024-25.pdf

Mishra, A., Das, N., & Chhetri, P. (2023). Sustainable strategies for the Indian coal sector: An econometric analysis approach. Sustainability, 15(14), 11129. https://doi.org/10.3390/su151411129

Mundlak, Y. (1978). On the pooling of time series and cross section data. Econometrica, 46(1), 69-85. https://doi.org/10.2307/1913646

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. CESifo Working Paper No. 1229; IZA Discussion Paper No. 1240. [Published version: Pesaran, M. H. (2021). General diagnostic tests for cross-sectional dependence in panels. Empirical Economics, 60(1), 13-50. https://doi.org/10.1007/s00181-020-01875-7]

Power Finance Corporation. (2024). Report on performance of state power utilities 2023-24. Power Finance Corporation Ltd.

REN21. (2024). Renewables 2024 global status report: Energy systems. REN21 Secretariat. https://www.ren21.net/gsr2024/modules/energy_systems_infrastructure/01_electricity_grids/

Shu, Y., Chen, G., He, J., & Zhang, F. (2020). Building a new electric power system based on new energy in China. Strategic Study of Chinese Academy of Engineering, 22(3), 61-69. https://doi.org/10.15302/J-SSCAE-2021.01.009

Singh, K., Pal, M., & Banerjee, R. (2024). Impact of institutional reforms on the performance of distribution utilities in India: A dynamic panel data analysis. Energy Policy, 186, 114001.

Verma, S., Agarwal, R., & Ghosh, S. (2020). Indian power distribution sector reforms: A critical review. Energy Policy, 143, 111418. https://doi.org/10.1016/j.enpol.2020.111418

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). MIT Press.

Appendix

Table A1. Dataset inventory: all sixteen source files used in the analysis

Dataset	Source Agency	Coverage	Frequency	Unit
Daily power generation	CEA	2017-2025	Daily	Station-unit
Daily power outages	CEA	2017-2025	Daily	Station-unit
Renewable generation (state)	CEA	2019/2020-2025	Daily	State
Daily coal stocks	CEA	2018-2025	Daily	Power station
Energy requirement/availability	CEA	Jan 2020-Jun 2023	Monthly	State
Installed capacity (state)	CEA	2017-2025	Monthly	State-sector
Interstate renewable generation	CEA	2019-2025	Daily	Station
State-controlled renewable gen.	CEA	2020-2025	Daily	State
Domestic coal production	MoC	2006-2023	Annual	State
Coal consumption (state)	MoC	2006-2021	Annual	State-sector
Coal imports	MoC	2006-2023	Annual	Country-type
Coal reserves by type	MoC	2006-2023	Annual	State
Coal reserves by depth	MoC	2006-2023	Annual	State
Coal reserves by field	MoC	2006-2023	Annual	State-field
Transmission lines commissioned	MoP	2014-2025	Project	National
Coal stocks at power stations	MoP	2018-2025	Daily	Power station

Note. CEA = Central Electricity Authority; MoC = Ministry of Coal; MoP = Ministry of Power. All datasets sourced from India's open data portal (data.gov.in) and the respective agency portals. Coverage windows determine the estimation sample as described in Section 3.3.