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.

Port Resilience During COVID-19: A Double Machine Learning and Equivalence Testing Approach

Do Huu Tam

2611414@student.hcmute.edu.vn

Ho Chi Minh City University of Technology and Engineering

Research Article

Keywords: Port resilience, sustainable development, liner shipping connectivity, COVID-19, double machine learning

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11598155/v1

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

Additional Declarations: No competing interests reported.

Port Resilience During COVID-19: A Double Machine Learning and Equivalence Testing Approach

Do Huu Tam[1*]

1Faculty of Economics, Ho Chi Minh City University of Technology and Engineering, Ho Chi Minh City, Vietnam.

*Corresponding author: 2611414@student.hcmute.edu.vn | ORCID: 0009-0003-2955-4592

Abstract

Maritime policy often assumes that connectivity, sustainability, and resilience reinforce one another, yet this assumption remains weakly tested. This study examines whether predetermined sustainable-development readiness strengthened the ability of liner connectivity to preserve container-port throughput during the COVID-19 shock. We combine the Sustainable Development Report 2026 backdated database with World Development Indicators for 149 maritime economies over 2004–2023. The main shock-response design compares 2020 with 2019 container throughput for 98 economies, including China and Viet Nam, using a readiness index that excludes infrastructure-related Goal 9 and controls for baseline throughput, pre-trends, income, trade exposure, manufacturing, internet use, and region. Inference combines HC3 OLS, repeated cross-fitted double/debiased machine learning, placebo and pre-trend tests, alternative indices, leave-one-country-out analysis, and equivalence testing. DML estimates are −0.0039 for readiness, 0.0353 for connectivity, and -0.0005 for their interaction. The interaction is statistically equivalent to zero within a policy-relevant ±0.05 log-point band. A secondary two-way fixed-effects model nevertheless shows a strong positive association between connectivity and throughput levels (0.5120; $p<0.001$). The evidence therefore distinguishes capacity from resilience: network connectivity supports scale, but does not by itself ensure shock absorption. Maritime policy should treat network expansion, sustainability performance, and operational resilience as related but non-substitutable policy domains, with separate metrics, investments, and accountability mechanisms.

Keywords: Port resilience; sustainable development; liner shipping connectivity; COVID-19; double machine learning

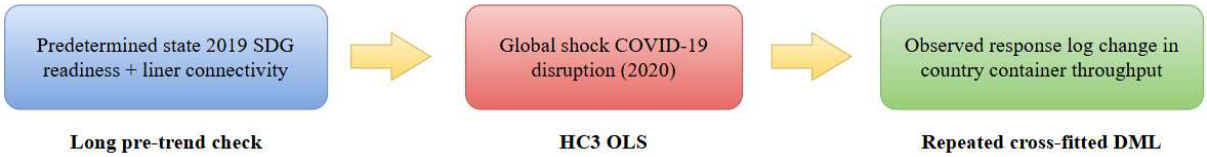

Figure 1. Identification logic

I. INTRODUCTION

The central decision for an aspiring maritime hub is not simply how to attract more services in normal times. It is how to preserve essential flows when many ports, carriers, workers, inland corridors, and customers are disrupted at once. Connectivity expands market access and supports scale, while sustainability can reduce externalities and strengthen institutional legitimacy. Resilience, however, is a performance property revealed under stress. Conflating the three creates a policy risk: governments may interpret a high connectivity index or a broad environmental social score as evidence that the port system can absorb, adapt to, and recover from disruption.

The empirical literature gives reasons for both optimism and caution. Liner connectivity is associated with trade access and lower transport frictions (Fugazza and Hoffmann, 2017), and port infrastructure can support seaborne trade and growth (Munim and Schramm, 2018). Port-sustainability research links operational, environmental, and social actions to the Sustainable Development Goals (SDGs), but remains dominated by conceptual frameworks, indicator development, and cases concentrated in leading ports (Lim et al., 2019; Alamoush et al., 2021). Recent resilience reviews emphasize robustness, redundancy, flexibility, rapidity, recoverability, and network reconfiguration rather than scale alone (Lau et al., 2024; Li et al., 2024). These literatures are adjacent, but their constructs are not interchangeable.

Earlier designs that regress bilateral trade on logistics and composite ESG indicators face five editor-level threats. First, the direction of causality between connectivity and trade is unresolved. Second, proprietary composite scores are difficult to reproduce. Third, excluding China materially limits external validity for ASEAN. Fourth, a short time dimension makes dynamic pooled mean-group or error-correction estimates vulnerable to small-T bias. Fifth, bilateral imports plus exports are an incomplete measure of hub performance because they do not distinguish transshipment, port handling, calls, dwell time, or re-export. The present study rebuilds the question around a direct maritime outcome and a common shock rather than adding complexity to the same weak identification structure. COVID-19 provides unusually clear timing for a global maritime disruption. UNCTAD documented widespread reductions in port calls and container activity in the first half of 2020, while AIS-based research estimated substantial changes in maritime mobility and trade (UNCTAD, 2020, 2022; Millefiori et al., 2021; Verschuur et al., 2021). The shock was not identical across countries, and baseline readiness was not randomly assigned. Nevertheless, fixing exposures in 2019, modelling the 2019–2020 change, conditioning on rich predetermined characteristics, verifying long pre-trends, and using flexible orthogonal estimation creates a more defensible shock-exposure comparison than a contemporaneous levels regression. It identifies a conditional resilience gradient not a universal structural effect of national sustainability.

The paper makes four contributions. First, it replaces a black-box ESG score with a reproducible national SDG-readiness measure constructed from goal-level data. Goal 9 is excluded from the primary index because it contains industry, innovation, and infrastructure content that could mechanically overlap with logistics. Second, it replaces bilateral trade with container-port throughput in twenty-foot equivalent units (TEU), a direct measure of maritime handling activity that includes transshipment lifts. Third, it combines conventional and machine-learning inference with tests designed for a credible null: repeated cross-fitting, long pre-trend checks, placebo years, alternative indices, China exclusion, leave-one-country-out estimates, multiple-testing control, and equivalence bounds. Fourth, it develops the capability-domain specificity proposition and translates it into a measurable resilience portfolio for Viet Nam and ASEAN.

The result is both negative and consequential. Neither national SDG readiness nor its interaction with liner connectivity buffered the 2020 throughput change in an economically material way. The DML interaction estimate is essentially zero, and its confidence interval lies within a ±0.05 log-point policy band. At the same time, connectivity is strongly associated with throughput levels in the longer panel. The appropriate conclusion is not that sustainability is unimportant. It is that broad sustainable-development performance is not a validated substitute for the operational capabilities that keep container systems running under synchronized disruption.

The remainder of this paper is structured as follows. Section 2 reviews the theoretical background and develops the research hypotheses. Section 3 outlines the data and measurement design, followed by the empirical strategy in Section 4. Section 5 presents the empirical results. Finally, Section 6 provides the discussion of findings, conclusions, and key implications.

II. THEORY AND HYPOTHESES

2.1. Capacity, sustainability, and resilience are different performance domains

A port system creates capacity through physical assets, service frequency, vessel size, carrier presence, direct country-pair links, and inland interfaces. LSCI captures several of these network-access attributes. High connectivity can increase the feasible scale and variety of trade, but may also concentrate dependence on tightly synchronized schedules, large vessels, dominant carriers, and congested interfaces. Capacity therefore expands the operating frontier without determining how much of that frontier remains available during disruption.

Sustainability is similarly multidimensional. Port actions range from emissions control and energy efficiency to labor standards, community relations, circularity, modal shift, and governance (Lam and Notteboom, 2014; Lu et al., 2016a, 2016b; Alamoush et al., 2021). A national SDG index captures the enabling environment surrounding these actions health, education, institutions, social inclusion, environmental quality, and economic conditions but it is not a direct audit of port-level practice. Its theoretical role here is readiness: a country with stronger broad systems may mobilize public services, information, labor protection, and coordination more effectively when a shock arrives.

Resilience is revealed by the time path of performance after disruption. In the engineering and supply-chain traditions, resilience combines resistance, absorption, adaptation, and recovery (Bruneau et al., 2003; Christopher and Peck, 2004). Dynamic-capabilities theory further suggests that sensing, seizing, and reconfiguring are context-specific routines rather than generic resource stocks (Teece et al., 1997). For ports, operational resilience may require spare gate and yard capacity, alternate labor teams, redundant power and data systems, diversified carrier and route portfolios, inland modal options, emergency customs procedures, and interoperable port-community data. These capabilities are much closer to the disruption mechanism than a national aggregate score.

2.2. Capability-domain specificity

The study proposes capability-domain specificity: complementarity should be expected only when the capability measured is sufficiently proximal to the performance domain and shock mechanism. Environmental quality and social development can improve long-horizon welfare and legitimacy without necessarily preventing a one-year drop in container lifts. Conversely, an emergency labor roster may protect throughput without improving a national sustainability index. Aggregation can therefore hide the relevant mechanism, and positive correlations among high-capacity, wealthy, and sustainable economies do not establish shock complementarity. *A broad capability index should not be presumed to amplify shock resilience unless its components map directly to the operational failure modes, substitution options, and recovery processes governing the outcome.*

The theory yields one capacity benchmark and two resilience hypotheses:

- ***H1*** *(capacity benchmark):* Liner connectivity is positively associated with container-throughput levels after country and year fixed effects and macro controls.
- ***H2*** *(readiness premium):* Higher predetermined SDG readiness is associated with a smaller container-throughput loss during the 2020 shock.
- ***H3*** *(complementarity):* Predetermined SDG readiness positively moderates the shock-resilience return to liner connectivity.

H1 is deliberately described as a benchmark association because changes in services and throughput can be jointly determined. H2 and H3 use a common-shock response design, but their causal interpretation remains conditional on exposure comparability after controls and pre-trend diagnostics. The equivalence analysis also tests the capability-domain proposition directly: if H3 is not only statistically insignificant but statistically smaller than a policy-relevant effect, broad SDG readiness cannot be defended as a material connectivity multiplier for this shock and outcome.

III. DATA AND MEASUREMENT

3.1. Data architecture and sample

The analysis merges two open, auditable databases by ISO-3 country code and year. The Sustainable Development Report (SDR) 2026 workbook provides backdated overall and goal-level scores from 2000 through 2025. The World Bank World Development Indicators API supplies container-port throughput, LSCI, real GDP and GDP per capita, population, trade openness, manufacturing value added, foreign direct investment, and internet use. The merged panel spans 2004–2023. A maritime-economy eligibility screen requires at least eight observed throughput years, eight LSCI years, and twelve SDG years, yielding 149 economies and 2,980 country-year records.

The main 2019–2020 shock specification uses complete cases for the outcome, both exposures, the pre-trend, and controls, producing 98 countries. China is retained. Viet Nam is retained. The broader diagnostics use up to 120 countries in pre-period years; the two-way fixed-effects benchmark covers 1,490 observations across 133 countries from 2010–2021, when LSCI coverage is strongest. Sample composition is reported rather than hidden behind a balanced-panel requirement, and every model-specific N appears in the replication workbook.

3.2. Outcome: direct container-port activity

Container-port traffic measures the flow of containers between land and sea modes in TEU. It includes coastal and international movements, empty units, and transshipment lifts; an intermediate transshipment is counted once when unloaded and again when reloaded. This makes TEU substantially closer to container-hub activity than bilateral trade value. The primary outcome is $\Delta \ln TEU_{i,2020} = \ln(TEU_{i,2020}) - \ln(TEU_{i,2019})$. It measures realized national handling performance during the first pandemic year, not an all-purpose latent hub status. Because annual country totals cannot reveal congestion, dwell time, service reliability, individual terminal performance, or value added, the paper consistently uses the terms container-port activity and throughput resilience.

3.3. Transparent sustainable-development readiness

The primary readiness score is an equal-weight arithmetic mean of the SDR backdated goal scores for Goals 1–8 and 10–17. Goal 9 is excluded ex ante because industry, innovation, and infrastructure can contain transport-adjacent content. Exclusion reduces the risk that logistics capacity is embedded in both the sustainability exposure and the maritime mechanism. All component scores, the aggregation rule, and the resulting country-year values are supplied. The score is therefore reproducible from public inputs rather than dependent on an unavailable proprietary transformation.

$$SDG\ readiness_{it} = (1/16)\ \Sigma g \in \{1\ldots,8,\ 10\ldots,17\}\ Goal\ score_{i}g_{t}$$

Two alternatives probe construct dependence: the published overall SDG index, which includes Goal 9, and an environment governance score averaging Goals 6, 7, 12, 13, 14, 15, and 16. The national scores are interpreted as contextual readiness, not direct port-level ESG performance. This distinction prevents a high national score from being mislabeled as evidence that a specific terminal has implemented resilient or green operating practices.

3.4. Connectivity and controls

LSCI captures the integration of a country into global container-shipping networks. Its components cover scheduled ships, deployed carrying capacity, maximum vessel size, services, companies, and under the updated methodology direct country-pair connections. The baseline exposure is ln(LSCI) in 2019. Both readiness and connectivity are standardized in the model sample before constructing their interaction. Controls are predetermined or measured at baseline: log 2019 throughput, the country-specific 2015–2019 slope in log throughput, log real GDP per capita, trade as a share of GDP, manufacturing value added as a share of GDP, internet use, and World Bank region indicators.

Table 1. Research design and measurement architecture

Element	Operationalization	Design role	Reviewer risk addressed
Container activity	Log change in national TEU, 2019–2020	Shock-performance outcome	Avoids trade-value/hub-status overreach
SDG readiness	Equal mean of Goals 1–8 and 10–17	Predetermined contextual exposure	Excludes infrastructure-rich Goal 9
Connectivity	Log 2019 country LSCI	Predetermined network-capacity exposure	Standardized before interaction
Pre-trend	Slope of log TEU, 2015–2019	Outcome-history control and diagnostic	Requires at least three observed years
Controls	Scale, income, trade, manufacturing, internet, region	Conditional comparability	No post-treatment controls in main model
Inference	HC3 OLS + repeated cross-fitted DML	Functional-form and overfitting robustness	Five folds; six repeated partitions
Meaningful null	TOST bound ±0.05 log points	Rules out policy-material effects	Approximately a five-percent throughput band

Table 2. Descriptive statistics for the 2020 complete-case sample

Variable	N	Mean	SD	Min	Median	Max
Δ log throughput, 2019–2020	98	-0.028	0.088	-0.454	-0.010	0.205
SDG readiness excluding Goal 9	98	71.033	8.396	52.725	71.985	86.532
Standardized log LSCI	98	0.446	0.848	-2.210	0.636	2.150
Log 2019 throughput	98	14.427	1.718	10.228	14.270	19.286
Pre-trend, 2015–2019	98	0.055	0.057	-0.103	0.043	0.265
Log real GDP per capita	98	9.193	1.250	6.425	9.271	11.289
Trade / GDP (%)	98	86.722	54.645	22.821	75.690	324.270
Manufacturing / GDP (%)	98	12.792	5.686	2.095	12.407	31.930
Internet users (%)	98	68.716	23.960	12.479	73.785	99.702

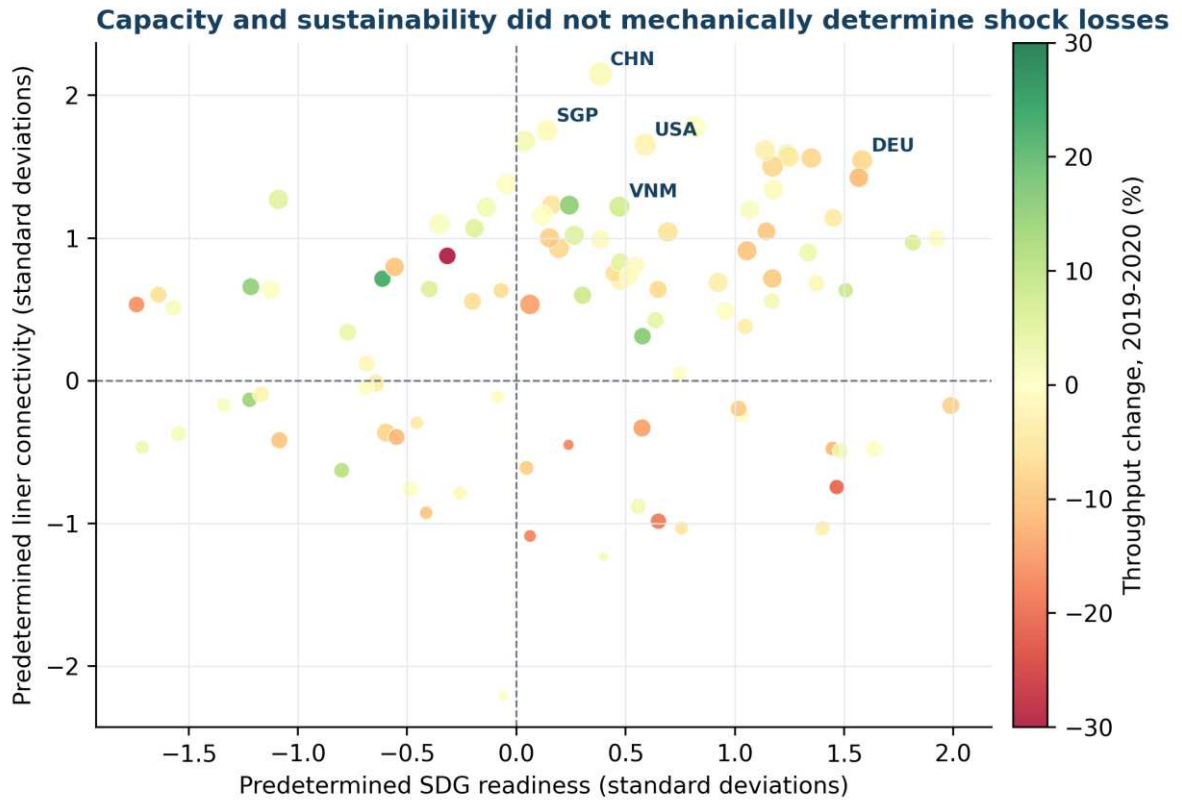

Figure 2. Predetermined capacity and SDG readiness did not mechanically determine 2020 throughput losses

IV. EMPIRICAL STRATEGY

4.1. Predetermined shock-exposure model

The main model relates the first-year throughput response to predetermined readiness and connectivity. By differencing log throughput relative to 2019 and controlling for baseline scale and pre-trend, the design removes time-invariant level differences from the outcome and reduces regression-to-the-mean risk. Region indicators absorb broad geography and institutional clusters. The estimating equation is:

$$\Delta lnTEU_{i,\,2020} = \alpha + \beta S\, S_{i,\,2019} + \beta C\, C_{i,\,2019} + \beta SC\, (S_{i,\,2019} \times C_{i,\,2019}) + \gamma' X_{i,\,2019} + \mu r + \varepsilon_i$$

S denotes standardized SDG readiness, C standardized log LSCI, X the control vector, and μr region fixed effects. βS is the conditional readiness gradient at average connectivity; βC is the conditional connectivity gradient at average readiness; βSC tests complementarity. The primary OLS specification uses HC3 heteroskedasticity-consistent standard errors, which are appropriate when leverage varies in a modest cross-section (MacKinnon and White, 1985). The outcome is winsorized at the sample 1st and 99th percentiles; raw-change results remain available in the workbook.

4.2. Repeated cross-fitted double machine learning

Linear controls can leave residual functional-form dependence when scale, income, openness, manufacturing, digital access, and region interact nonlinearly. DML estimates a low-dimensional treatment vector while machine-learning nuisance functions absorb flexible relations with X (Chernozhukov et al., 2018). Extra Trees predicts the outcome and random forests predict each treatment component. Five-fold cross-fitting ensures that every residual is generated out of fold. Six independently shuffled partitions are averaged to reduce sensitivity to a single split. The final orthogonal regression uses HC3 inference on residualized outcomes and exposures.

$$Y_i = m(X_i) + \theta' D_i + u_i;\; D_i = g(X_i) + v_i;\; E[u_i v_i] = 0$$

D contains readiness, connectivity, and their product. This partially linear parameter is causal only if the conditioning set is sufficient for exposure comparability and there are no unmeasured differential shocks correlated with both the exposures and throughput response. DML protects against regularization and nuisance-model bias;

it does not create random assignment. The paper therefore describes the result as orthogonal shock-exposure evidence and keeps the identifying assumptions visible.

4.3. Diagnostics, equivalence, and secondary panel benchmark

Three diagnostic layers address the difference between 'not significant' and 'not meaningful.' First, year-specific changes relative to 2019 are estimated for 2014–2023. Pre-2019 coefficients function as negative-control years, and Benjamini–Hochberg q-values limit false discovery across horizons (Benjamini and Hochberg, 1995). A separate diagnostic regresses the estimated 2015–2019 throughput slope on the exposures and controls. Second, leave-one-country-out estimates test whether China, Singapore, the United States, small islands, or other influential observations drive the result. Third, TOST evaluates whether the SDG and interaction effects lie inside ±0.05 log points. This smallest effect of policy interest corresponds approximately to a five-percent throughput difference for a one-standard-deviation exposure change. Equivalence is concluded only if both one-sided null hypotheses are rejected (Lakens, 2017).

The secondary capacity benchmark estimates log throughput levels over 2010–2021 with country and year fixed effects and country-clustered standard errors. It includes time-varying standardized readiness, log LSCI, their interaction, log real GDP per capita, and trade openness. Country effects absorb time-invariant geography, institutional history, and long-run port endowment; year effects absorb global cycles. The model is not used to claim that connectivity causes scale, because service deployment and throughput remain jointly determined. Its purpose is to show whether the capacity relationship differs from the shock-resilience relationship.

V. RESULTS

The 98-country main sample experienced a mean 2019–2020 change of -0.028 log points and a median of -0.010. Dispersion was substantial: the minimum was -0.454 and the maximum 0.205. This heterogeneity creates meaningful variation to explain, but Figure 2 shows no simple quadrant in which high sustainability and high connectivity guarantee positive performance. High-capacity economies appear on both sides of zero, as do economies with above-average readiness.

Table 3. Predetermined exposure estimates for the 2019–2020 throughput response.

Term	OLS-HC3	p	DML-HC3	p	DML 95% CI	Equivalent ±0.05
SDG readiness	-0.0210 (0.0192)	0.277	-0.0039 (0.0211)	0.853	[-0.0453, 0.0374]	Yes
Liner connectivity	0.0446 (0.0312)	0.157	0.0353* (0.0207)	0.091	[-0.0052, 0.0759]	No
SDG x connectivity	-0.0002 (0.0141)	0.991	-0.0005 (0.0140)	0.970	[-0.0280, 0.0270]	Yes

The OLS readiness coefficient is −0.0210 (SE 0.0192; p=0.277), connectivity is 0.0446 (0.0312; p=0.157), and the interaction is −0.00015 (0.0141; p=0.991). Flexible orthogonalization moves readiness closer to zero: the DML estimate is −0.0039 (0.0211; p=0.853). Connectivity is positive but imprecise at 0.0353 (0.0207; p=0.091). Most importantly, the DML interaction is −0.0005 (0.0140; p=0.970). The sign and magnitude provide no support for the claim that broad SDG readiness amplified the 2020 shock return to liner connectivity. H2 and H3 are not supported.

Equivalence inference sharpens the conclusion. The DML readiness confidence interval is [−0.0453, 0.0374], and TOST rejects effects outside ±0.05 (p=0.0145). The interaction interval is [−0.0280, 0.0270], with TOST p=0.00021. Accordingly, the data do more than fail to reject zero: under the stated design they rule out a readiness main effect and connectivity multiplier larger than the declared policy threshold. Connectivity itself is neither statistically different from zero at five percent nor equivalent to zero because its upper confidence limit remains economically relevant.

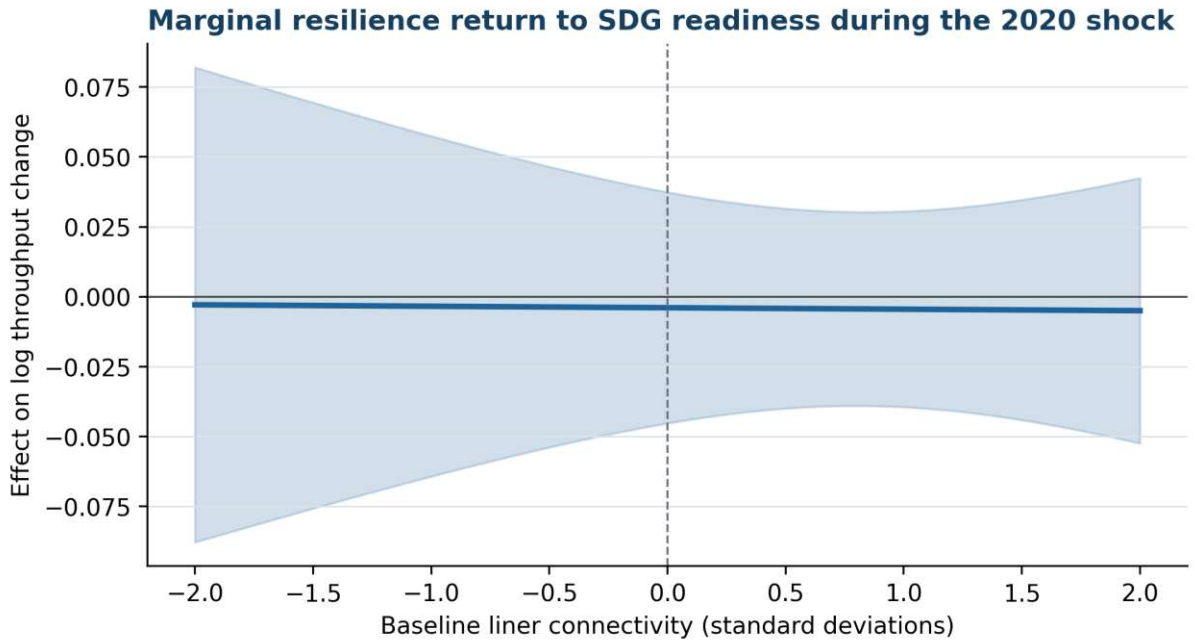

Figure 3. The marginal shock-response return to SDG readiness is flat across baseline connectivity

Table 4. Secondary two-way fixed-effects benchmark for log throughput levels, 2010–2021

Term	Coefficient	Clustered SE	p
SDG readiness	0.0588	0.1422	0.6798
Liner connectivity	0.5120***	0.1147	0.0000
SDG × connectivity	-0.0327	0.0536	0.5422
Log real GDP per capita	0.7345***	0.2067	0.0005
Trade/GDP	0.0335	0.0417	0.4230

The fixed-effects benchmark supports H1 as an association: a one-standard-deviation increase in log LSCI is associated with a 0.5120 increase in log throughput (SE 0.1147; $p<0.001$). The SDG coefficient is 0.0588 (0.1422; $p=0.680$), and the interaction is −0.0327 (0.0536; $p=0.542$). Thus connectivity tracks within-country capacity and activity over time, while neither the broad readiness score nor complementarity is detected. The contrast with Table 3 is the paper's central empirical separation: a network attribute that covaries strongly with scale does not automatically stabilize performance during a synchronized shock.

Table 5. Robustness and negative-control estimates.

Specification	SDG	p	Interaction	p	N
Main: SDG ex-infrastructure	-0.0210	0.277	-0.0002	0.991	98
Overall SDG index	-0.0197	0.357	-0.0028	0.834	98
Environmental-governance SDG	-0.0154	0.200	0.0085	0.513	98
Exclude China	-0.0210	0.277	-0.0001	0.993	97
Population >= 1 million	-0.0134	0.561	0.0071	0.652	91
2021 recovery	-0.0039	0.916	-0.0073	0.785	95
Placebo 2018	0.0094	0.746	0.0350	0.084	120
Placebo 2017	0.0315	0.375	0.0303	0.337	118

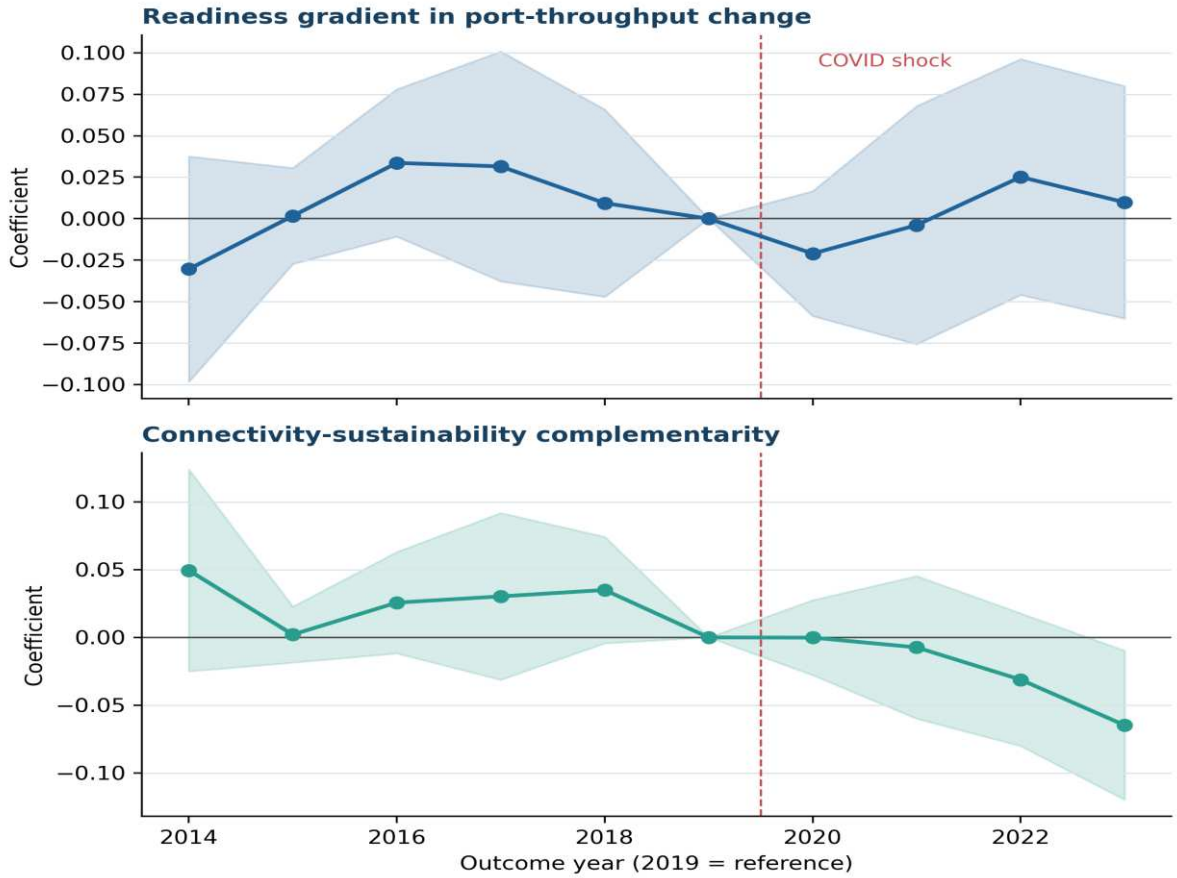

Figure 4. Year-specific readiness and complementarity gradients relative to 2019.

Results are insensitive to index definition and sample composition. The readiness estimate remains negative when the overall SDG index or environment governance index replaces the primary score, excluding China changes the fourth decimal place, and the population threshold does not recover a positive effect. The 2021 recovery outcome is also null. Leave-one-country-out estimates are especially informative for the interaction: across 98 re-estimations, the coefficient ranges only from -0.0059 to 0.0029. The readiness range is -0.0282 to -0.0116; connectivity remains positive in every omission, from 0.0340 to 0.0612.

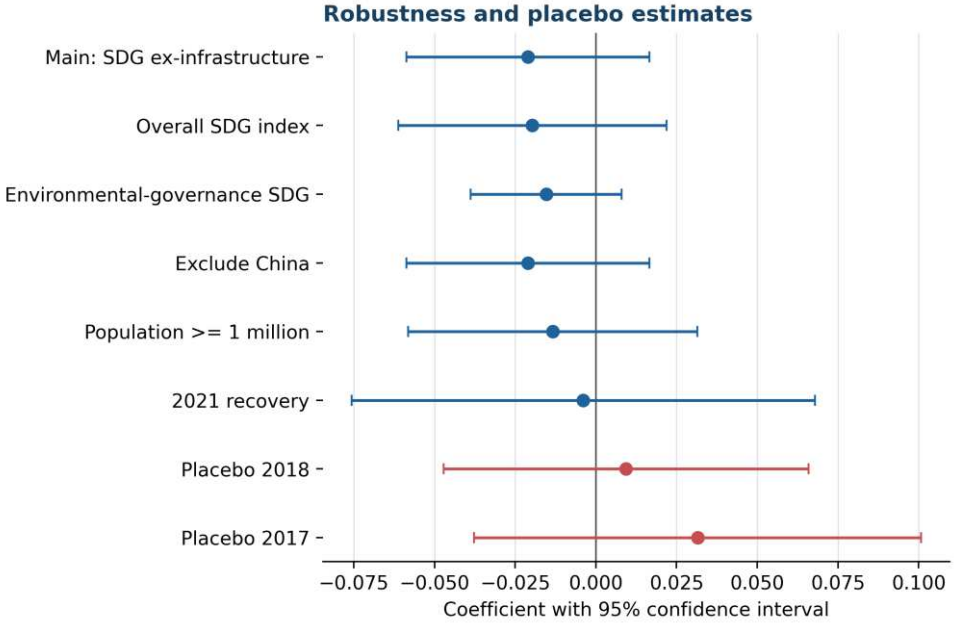

Figure 5. Readiness estimates remain small across alternative measures, samples, recovery, and placebo years

VI. DISCUSSION

6.1. Discussion of research results

A conventional null would leave two interpretations: no true effect or insufficient precision. The equivalence result materially narrows that ambiguity. For the chosen outcome, shock, sample, conditioning set, and ±0.05 policy threshold, a large SDG connectivity complementarity is inconsistent with the data. This is not a claim that every dimension of sustainable development has zero operational value. It is evidence against using a broad national composite as if it were a validated port-resilience multiplier.

The result also clarifies why previous positive levels associations can be misleading. High-income countries often have better institutions, sustainability outcomes, connectivity, and port scale simultaneously. A pooled model can translate this bundle into a positive coefficient even if the broad score does not protect throughput when stress arrives. Country and year effects reduce but do not eliminate the endogeneity of connectivity; the shock-response design changes the estimand from 'who has large ports?' to 'who changed less in the same year?' The difference in estimands explains why the capacity benchmark is strong while the resilience gradient is weak.

The sustainability–resilience literature often assumes a general halo: an economy or port that performs well on environmental and social dimensions should also be more robust. Capability-domain specificity places a boundary on that logic. Broad systems matter when they activate mechanisms relevant to the disruption. Health systems, digital public services, worker protection, and effective institutions may support continuity; emissions outcomes or biodiversity protection may not immediately increase container lifts. Aggregating them can average away the operational signal.

For theory, this shifts the question from whether sustainability and resilience are complementary in the abstract to which sustainability capability, against which shock, through which operational mechanism, at what time horizon. It also separates stock from response. Connectivity is a network stock, SDG readiness a broad contextual stock, and resilience a realized response path. Complementarity should be tested at the response level with domain-proximate measures such as route concentration, spare capacity, digital redundancy, labor continuity, inland modal substitution, emergency customs performance, and time to recovery.

The findings are compatible with research showing that connectivity supports trade and scale, while extending it by rejecting the inference that scale equals resilience. They also answer calls in port-sustainability reviews for empirical tests using realized performance rather than only frameworks and proposed actions (Lim et al., 2019; Alamoush et al., 2021). Recent resilience reviews classify redundancy, flexibility, and recoverability as central attributes (Lau et al., 2024; Li et al., 2024). The present evidence explains why those attributes should be measured directly: national SDG readiness does not recover them reliably. COVID-19 port studies document heterogeneous losses and recovery paths, frequently using AIS, network, case-study, or composite-resilience approaches. The country-level TEU design complements rather than replaces that work. It trades operational granularity for global coverage, transparent replication, and a direct test of a policy narrative. The resulting null should therefore be interpreted as a construct warning: broad national readiness is too distal to stand in for terminal and corridor continuity.

6.2. Policy and managerial implications

The immediate policy implication is not to deprioritize sustainability. It is to stop asking one composite score to perform three jobs. Port authorities and ministries should maintain separate but coordinated ledgers for capacity, sustainability, and resilience. Each ledger needs an outcome, an accountable owner, and a stress-tested target.

Table 6. Three-ledger governance for a sustainable and resilient maritime hub

Ledger	Core metrics	Decision objective	Failure signal
Network capacity	Direct services; deployed TEU; carrier and route diversity; maximum vessel; hinterland capacity	Reliable access and scalable volume	Service loss; concentration; berth/yard saturation
Sustainability	GHG intensity; shore-power use; renewable share; safety; community impact; modal split	Lower externalities and durable license to operate	Carbon lock-in; local harm; regulatory exposure
Operational resilience	Time to recover; spare gate/yard; alternate crews; power/data backup; inland substitution	Maintain and restore critical flows	Single points of failure; slow recovery; cascading queues

Investment appraisal should therefore attach two additional tests to conventional net present value and carbon analysis. A dependency test identifies single points of failure across vessel services, tugs, pilots, cranes, gates, customs, electricity, data, labor, roads, rail, and barges. A recovery test specifies the throughput level that must be maintained, the maximum tolerable interruption, and the resources required to restore service. Green investments can strengthen resilience when they remove a failure mode for example, distributed energy with islanding capability or digital systems with offline continuity but the mechanism must be demonstrated rather than inferred from an ESG label.

Viet Nam's national container throughput increased from 15.30 million TEU in 2019 to 16.39 million in 2020, a 7.18% gain in the source data. This observed outcome is notable but should not be treated as proof that national SDG readiness or connectivity caused resilience. It may reflect trade composition, production relocation, timing, port mix, policy response, and partner demand. The correct managerial response is to convert a favorable realized outcome into designed, measurable continuity rather than assume the system is inherently protected.

1. Diversify deep-sea access and monitor concentration. Track service, alliance, carrier, and destination concentration at Cai Mep–Thi Vai and other gateways; define trigger levels for loss of a major service or carrier.
2. Build inland evacuation redundancy. Treat barge, road, rail, ICD, and dry-port capacity as a substitution portfolio. Measure how quickly containers can be rerouted when a gate, bridge, channel, or terminal is constrained.
3. Create a continuity-grade port community system. Interoperability should include customs, terminal, carrier, trucking, depot, and inland modes, with cybersecurity segmentation, offline procedures, data backup, and tested restoration time.
4. Institutionalize labor and contractor continuity. Maintain cross-trained teams, mutual-aid agreements, health and safety protocols, and minimum staffing plans for pilots, tugs, cranes, gates, customs, and maintenance.
5. Link decarbonization to failure-mode removal. Shore power, electrification, renewable energy, and efficiency projects should report emissions effects and continuity effects separately, including backup duration, fuel or energy diversity, and recovery time.
6. Publish a port-level resilience dashboard. Annual national TEU is too coarse for management. Monthly calls, omitted sailings, berth and anchorage time, dwell, yard density, gate time, rail/barge share, re-export, transshipment, and time-to-recover should be available by port and terminal where commercial confidentiality permits.

ASEAN economies entered the same global shock with sharply different outcomes. Among observed members, 2019–2020 throughput changes ranged from −14.08% in the Philippines to +7.18% in Viet Nam; Singapore declined 0.87%, Malaysia increased 1.73%, and Indonesia declined 7.35%. The dispersion confirms that regional membership, connectivity, or broad readiness alone does not determine realized continuity.

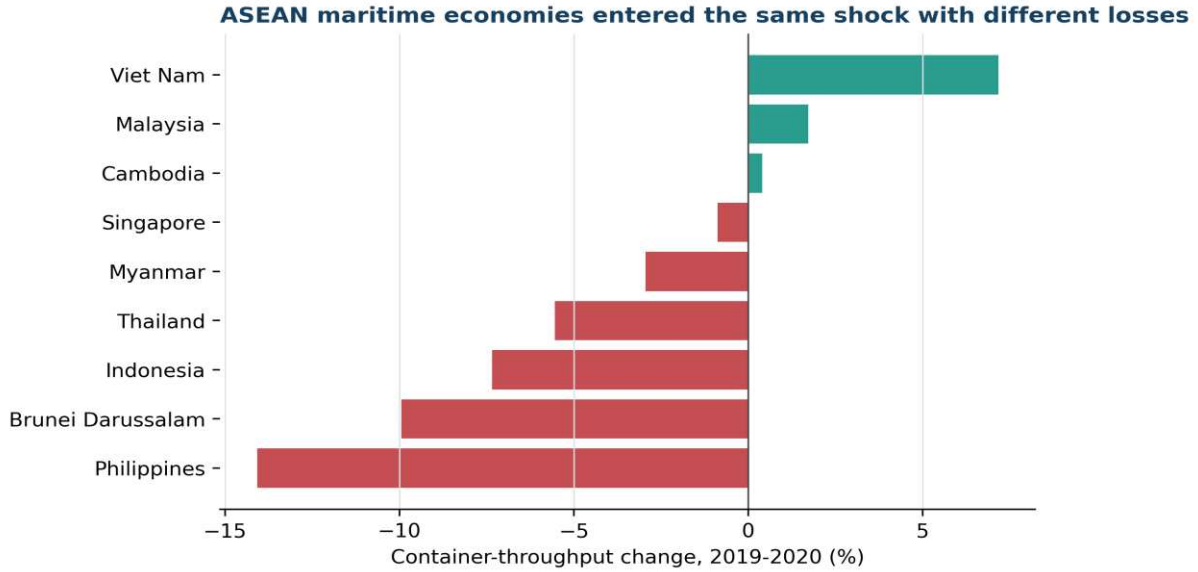

Figure 6. ASEAN maritime economies experienced markedly different 2019–2020 container-throughput changes

Regional resilience policy should focus on interoperability and substitution. Priorities include common port-call and disruption data standards; emergency customs and sanitary protocols; mutual recognition of electronic documents; corridor-level contingency exercises; visibility of spare gateway and feeder capacity; and pre-arranged alternatives when a hub, strait, or land corridor is constrained. ASEAN can also standardize a compact resilience scorecard based on observable recovery outcomes rather than another composite of distant national attributes.

6.3. Robustness boundaries and limitations

The design materially improves causal discipline but does not eliminate all identification risk. COVID-19 provides common timing, not identical treatment intensity. Infection waves, containment policy, sector mix, stimulus, port governance, carrier decisions, and trading-partner demand varied across economies. Baseline controls and long pre-trends reduce confounding; DML reduces functional-form sensitivity; neither substitutes for random assignment or a credible instrument for readiness. The estimand is therefore conditional and shock-specific.

Measurement is direct but aggregated. National TEU counts transshipment and is closer to hub activity than bilateral trade, yet it cannot reveal which ports or terminals changed, whether vessels waited longer, whether cargo value or reliability changed, or how much re-export occurred. Resilience is observed as annual throughput preservation and subsequent recovery, not as a complete multidimensional index. Monthly port-level throughput, AIS calls, dwell, reliability, crane and gate performance, and inland flows would support stronger mechanism tests.

The SDG database is transparent but remains a constructed national index. Backdating, normalization, missing-data procedures, and changes in source indicators can introduce measurement error. Excluding Goal 9 prevents direct infrastructure overlap but also removes some genuinely relevant innovation and digital capacity. Alternative index checks reduce, but do not remove, this construct trade-off. The result should not be generalized to verified port-level sustainability practices such as energy redundancy, worker protection, digital governance, or modal diversification.

The complete-case shock sample contains 98 countries and one global disruption. China and ASEAN are present, but missingness may still be selective, particularly for small and low-capacity economies. The 2010–2021 fixed-effects benchmark has a longer time dimension than the original short panel, yet connectivity remains endogenous and its measurement methodology evolved. Future work should preregister shock-specific mechanisms, use monthly port microdata, exploit staggered natural disruptions or infrastructure openings, and test multiple shock types pandemics, cyber incidents, drought, extreme weather, strikes, chokepoint closure, and geopolitical rerouting.

VII. CONCLUSION

This study rebuilds the sustainability connectivity hub question around a direct outcome, an auditable exposure, a common shock, and inference capable of supporting a meaningful null. In 98 maritime economies, predetermined national SDG readiness neither buffered the 2020 container-throughput change nor amplified the return to liner connectivity. Repeated cross-fitted DML places both estimates near zero, and equivalence testing rules out an interaction larger than ±0.05 log points. China's inclusion does not change the result. Alternative SDG measures, population thresholds, placebo years, pre-trend tests, and leave-one-country-out estimates lead to the same conclusion.

Connectivity nonetheless remains strongly associated with throughput levels in the longer two-way fixed-effects panel. The combined evidence establishes the paper's central distinction: capacity is not resilience. A country can be highly connected and still exposed to synchronized disruption; a country can perform well on broad sustainable-development indicators without possessing terminal, route, inland, labor, power, and digital redundancy.

For policy, the implication is precise. Do not sequence connectivity 'before' sustainability, and do not treat sustainability as a resilience proxy. Govern three ledgers capacity, sustainability, and operational resilience with distinct metrics and owners, then invest where mechanisms overlap. For Viet Nam and ASEAN, the highest-value next step is a port-level continuity data and stress-testing system that converts network expansion and green transition into demonstrable recovery capability.

Funding. The author received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests. The author declares no competing interests.

Ethics approval. Not applicable. The study uses aggregated, publicly available country-level data and involves no human participants or personal data.

Data availability. The cleaned country-year panel, model outputs, data dictionary, exact source URLs, and an auditable replication workbook accompany this manuscript.

Author contributions. Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing original draft, writing review and editing, visualization, and project administration.

References

Alamoush, A. S., Ballini, F., & Ölçer, A. I. (2021). Revisiting port sustainability as a foundation for the implementation of the United Nations Sustainable Development Goals (UN SDGs). Journal of Shipping and Trade, 6, 19. https://doi.org/10.1186/s41072-021-00101-6

Angrist, J. D., & Pischke, J.-S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 57(1), 289–300.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., & von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. Earthquake Spectra, 19(4), 733–752. https://doi.org/10.1193/1.1623497

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21(1), C1–C68. https://doi.org/10.1111/ectj.12097

Christopher, M., & Peck, H. (2004). Building the resilient supply chain. The International Journal of Logistics Management, 15(2), 1–14. https://doi.org/10.1108/09574090410700275

Davarzani, H., Fahimnia, B., Bell, M., & Sarkis, J. (2016). Greening ports and maritime logistics: A review. Transportation Research Part D: Transport and Environment, 48, 473–487. https://doi.org/10.1016/j.trd.2015.07.007

Fugazza, M., & Hoffmann, J. (2017). Liner shipping connectivity as determinant of trade. Journal of Shipping and Trade, 2, 1. https://doi.org/10.1186/s41072-017-0019-5

Holling, C. S. (1973). Resilience and stability of ecological systems. Annual Review of Ecology and Systematics, 4, 1–23.

Imbens, G. W., & Rubin, D. B. (2015). Causal Inference for Statistics, Social, and Biomedical Sciences. Cambridge University Press.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. Social Psychological and Personality Science, 8(4), 355–362. https://doi.org/10.1177/1948550617697177

Lam, J. S. L., & Notteboom, T. (2014). The greening of ports: A comparison of port management tools used by leading ports in Asia and Europe. Transport Reviews, 34(2), 169–189. https://doi.org/10.1080/01441647.2014.891162

Lau, Y.-Y., Chen, Q., Poo, M. C. P., Ng, A. K. Y., & Ying, C. C. (2024). Maritime transport resilience: A systematic literature review on the current state of the art, research agenda and future research directions. Ocean & Coastal Management, 251, 107086. https://doi.org/10.1016/j.ocecoaman.2024.107086

Laxe, F. G., Bermúdez, F. M., Palmero, F. M., & Novo-Corti, I. (2017). Assessment of port sustainability through synthetic indexes: Application to the Spanish case. Marine Pollution Bulletin, 119(1), 220–225. https://doi.org/10.1016/j.marpolbul.2017.03.064

Li, K. X., Wang, H., Yang, Y., & Li, M. (2024). Resilience in maritime logistics: Theoretical framework, research methodology, and indicator system. Ocean & Coastal Management, 259, 107465. https://doi.org/10.1016/j.ocecoaman.2024.107465

Li, W., Asadabadi, A., & Miller-Hooks, E. (2022). Enhancing resilience through port coalitions in maritime freight networks. Transportation Research Part A: Policy and Practice, 157, 1–23. https://doi.org/10.1016/j.tra.2022.01.015

Lim, S., Pettit, S., Abouarghoub, W., & Beresford, A. (2019). Port sustainability and performance: A systematic literature review. Transportation Research Part D: Transport and Environment, 72, 47–64. https://doi.org/10.1016/j.trd.2019.04.009

Lu, C.-S., Lai, P.-L., & Chiang, Y.-P. (2016). Container terminal employees' perceptions of the effects of sustainable supply chain management on sustainability performance. Maritime Policy & Management, 43(5), 597–613. https://doi.org/10.1080/03088839.2016.1190471

Lu, C.-S., Shang, K.-C., & Lin, C.-C. (2016). Examining sustainability performance at ports: Port managers' perspectives on developing sustainable supply chains. Maritime Policy & Management, 43(8), 909–927. https://doi.org/10.1080/03088839.2016.1199918

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics, 29(3), 305–325. https://doi.org/10.1016/0304-4076(85)90158-7

Millefiori, L. M., Braca, P., Zissis, D., Spiliopoulos, G., Marano, S., Willett, P. K., & Carniel, S. (2021). COVID-19 impact on global maritime mobility. Scientific Reports, 11, 18039. https://doi.org/10.1038/s41598-021-97461-7

Munim, Z. H., & Schramm, H.-J. (2018). The impacts of port infrastructure and logistics performance on economic growth: The mediating role of seaborne trade. Journal of Shipping and Trade, 3, 1. https://doi.org/10.1186/s41072-018-0027-0

Nguyen, T.-T., Tran, D., Ton, H. D. T., & Thai, V. V. (2023). Managing disruptions in the maritime industry: A systematic literature review. Maritime Business Review, 8(2), 170–190. https://doi.org/10.1108/MABR-09-2021-0072

Notteboom, T. E., & Rodrigue, J.-P. (2005). Port regionalization: Towards a new phase in port development. Maritime Policy & Management, 32(3), 297–313. https://doi.org/10.1080/03088830500139885

Sachs, J. D., Lafortune, G., Fuller, G., & Iablonovski, G. (2026). Implementing Sustainable Development: 2030 and Beyond. Sustainable Development Report 2026. SDSN and Dublin University Press. https://doi.org/10.82163/225

Sheffi, Y., & Rice, J. B. (2005). A supply chain view of the resilient enterprise. MIT Sloan Management Review, 47(1), 41–48.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. Strategic Management Journal, 18(7), 509–533.

Tsoulfas, G. T. (2025). Port resilience: A systematic literature review. Maritime Economics & Logistics. Advance online publication. https://doi.org/10.1057/s41278-025-00326-3

United Nations Conference on Trade and Development. (2020). Review of Maritime Transport 2020. United Nations.

United Nations Conference on Trade and Development. (2022). COVID-19 and Maritime Transport: Impact and Responses. United Nations.

Verschuur, J., Koks, E. E., & Hall, J. W. (2021). Global economic impacts of COVID-19 lockdown measures stand out in high-frequency shipping data. PLOS ONE, 16(4), e0248818. https://doi.org/10.1371/journal.pone.0248818

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48(4), 817–838.

World Bank. (2026). World Development Indicators [Data set]. https://databank.worldbank.org/source/world-development-indicators

Appendix A. Exact variable lineage and model inventory

Table A1. Exact source-variable lineage.

Series / sheet	Provider	Definition	Use
IS.SHP.GOOD.TU	World Bank WDI / UNCTAD	Container port traffic (TEU)	Outcome and panel levels
IS.SHP.GCNW.XQ	World Bank WDI / UNCTAD	Liner Shipping Connectivity Index	Baseline and panel connectivity
NY.GDP.PCAP.KD	World Bank WDI	Real GDP per capita	Baseline and panel control
NE.TRD.GNFS.ZS	World Bank WDI	Trade openness	Baseline and panel control
NV.IND.MANF.ZS	World Bank WDI	Manufacturing value added	Baseline control

IT.NET.USER.ZS	World Bank WDI	Internet users	Baseline control
Backdated SDG Index	SDSN SDR 2026	Overall and Goal 1–17 scores	Primary and alternative readiness