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.

Behind the income gap in the environmental efficiency of national aviation systems: a metafrontier analysis with bootstrap inference

Quynh Nga DUONG

ngadq@vaa.edu.vn

Vietnam Aerospace University

Article

Keywords: Aviation, Environmental efficiency, Data envelopment analysis, Metafrontier, Technology gap, Bootstrap, Global Malmquist–Luenberger index

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11396829/v1

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

Additional Declarations: No competing interests reported.

Behind the income gap in the environmental efficiency of national aviation systems: a metafrontier analysis with bootstrap inference

Quynh Nga DUONG
Vietnam Aerospace University, Ho Chi Minh City, Vietnam
ORCID: 0000-0002-1392-9201
Corresponding author: Quynh Nga DUONG — Email: ngadq@vaa.edu.vn

Abstract

Single-frontier comparisons often rank the aviation systems of lower-income countries as environmentally inefficient without showing why. This study measures the environmental efficiency of national air transport in 84 countries over 2010–2023 with a slacks-based data envelopment analysis model, an income-group metafrontier and a country-block bootstrap, treating transport-sector CO_2 as an undesirable output. Efficiency against the metafrontier falls with income, but efficiency against each group's own frontier does not. The gradient lies in the technology gap ratio (0.939, 0.462 and 0.165 for high-, upper-middle- and lower-middle- and low-income groups), and the gap between the high-income group and each other group is significant. Freight output accounts for most measured inefficiency, and five hub economies dominate the frontier. Point estimates indicate productivity growth of about 2% a year, driven by shifts in best practice rather than catch-up. The results suggest that benchmarking should report within-group performance and technology gaps separately.

Keywords: Aviation; Environmental efficiency; Data envelopment analysis; Metafrontier; Technology gap; Bootstrap; Global Malmquist–Luenberger index

Introduction

Aviation is a persistent and growing source of carbon dioxide emissions. Civil aviation emitted about 915 million tonnes of CO_2 in 2019, a little over 2% of anthropogenic emissions, and its demand is growing fastest in developing countries (Queiroz Júnior et al., 2024). Because the sector's output is valued in passengers and tonne-kilometres rather than in value added, a useful question for policy is not only how much a country's air transport emits, but how much transport service it delivers per unit of infrastructure, flights and emissions. Data envelopment analysis (DEA) addresses this question by comparing countries against a best-practice frontier without imposing a functional form or prices on inputs and outputs (Charnes et al., 1978; Banker et al., 1984).

Country-level evidence is limited. Most DEA studies of aviation evaluate airlines or airports (Yakath Ali et al., 2021), and the few that evaluate national systems, such as Queiroz Júnior et al. (2024), use small samples, include macroeconomic aggregates in the production technology and report point estimates without measures of uncertainty. They also compare all countries against a single frontier, which assumes that any national system could reach the best practice observed anywhere. Countries at different income levels differ in fleets, airport networks and traffic mix, so a single frontier may confound inefficiency with differences in the technology available to a country's aviation system.

This study measures the environmental efficiency and productivity of national aviation systems in 84 countries over 2010–2023 and asks three questions. (i) Do national aviation systems at different income levels differ in environmental efficiency, and if so, is the difference due to inefficiency relative to comparable systems or to a gap between the best practice available to each income group and the overall best practice? (ii) How did productivity change over the period, and did it come from catch-up, from shifts in best practice within groups or from changes in the technology gap? (iii) Which country conditions are associated with efficiency once the estimation uncertainty of the scores is taken into account?

To answer these questions, the study combines established methods. A slacks-based measure (SBM) with CO_2 as an undesirable output (Tone, 2004) is estimated against a metafrontier and three income-group frontiers (O'Donnell et al., 2008), and productivity change is decomposed with the metafrontier decomposition of Oh and Lee (2010) applied to the global Malmquist–Luenberger index with undesirable outputs (Oh, 2010). All inference uses a country-block bootstrap in which every frontier is re-estimated in every replicate, so that the intervals for scores, technology gaps, productivity components and second-stage coefficients reflect the estimation of the frontiers. Macroeconomic aggregates are kept out of the production technology, the sample is defined by a size rule fixed before estimation, and the results are re-estimated without the economies that host internationally based carriers and in a 155-economy sample.

The paper makes three contributions. First, it provides country-level evidence for a larger and more diverse sample than earlier work and separates inefficiency within income groups from the technology gap between them. Second, it attaches bootstrap intervals to the scores, technology gap ratios, productivity components and second-stage coefficients, so that differences between groups can be assessed statistically. Third, it shows how two features of

the data shape the results: the concentration of the frontier on economies that host international carriers and the dominance of freight in the measured inefficiency. The remainder of the paper reviews the related literature, describes the data and methods, presents and discusses the results, and concludes.

Related literature

DEA with undesirable outputs in aviation

Efficiency studies of aviation have mostly used airlines or airports as decision-making units. Arjomandi and Seufert (2014) used bootstrapped DEA under variable returns to scale to assess the technical and environmental efficiency of 48 major airlines over 2007–2010, and Huang et al. (2020) evaluated the productivity of 15 international airlines over 2011–2017 with a modified Malmquist model that penalised both CO_2 emissions and flight delays. Kaya Aydin and Aydin (2022) found no convergence in the CO_2 efficiency of airlines over 2011–2018. A review of four decades of airline efficiency studies shows that DEA and its Malmquist extensions remain the dominant tools and that environmental variables have been added mainly in the last decade (Yakath Ali et al., 2021). Several recent studies examine the pandemic: Voltes-Dorta et al. (2024) found a productivity loss of 10–14% for 34 airline groups over 2019–2022, Wu et al. (2024) found that the efficiency of 26 airlines fell sharply in 2020 and had only partly recovered by 2022, and Li et al. (2022) reported a similar fall in the environmental efficiency of 14 airlines. At the airport level, Ngo et al. (2026) combined an SBM model with undesirable outputs and a truncated second-stage regression for Vietnamese airports.

Two modelling choices recur in this literature. Directional distance functions expand good outputs and contract bad outputs along a chosen direction (Chung et al., 1997), whereas the slacks-based measure (SBM) of Tone (2001), extended to undesirable outputs by Tone (2004), uses the input and output slacks directly and yields a non-radial score between 0 and 1. For change over time, the global Malmquist index of Pastor and Lovell (2005), extended by Oh (2010) to undesirable outputs, uses a single frontier formed by all periods, which guarantees feasibility and makes the index circular.

National aviation systems and technological heterogeneity

Evidence at the level of national aviation systems is thin. Queiroz Júnior et al. (2024) applied an output-oriented BCC model to 21 countries between 2008 and 2018, with airports and population as inputs, passengers, cargo, departures and GDP as outputs, and transport CO_2 as

the undesirable output. Ten countries were fully efficient in all years, including the United States, China, India and Germany. Placing GDP and population inside the technology can favour large economies, and the small sample and the absence of measures of uncertainty make it difficult to know which differences are systematic.

A single frontier for all countries assumes that every national system could, in principle, reach the best practice observed anywhere in the sample. Countries at different income levels differ in fleet composition, airport infrastructure, network structure and the share of long-haul and cargo operations, so this assumption may not hold. The metafrontier approach addresses such heterogeneity by estimating separate frontiers for groups of units and an envelope of all of them (Battese et al., 2004; O'Donnell et al., 2008). The ratio of a unit's efficiency measured against the metafrontier to its efficiency measured against its group frontier, the technology gap ratio, separates inefficiency relative to the unit's own group from the distance between the group's best practice and the overall best practice. Oh and Lee (2010) combined the metafrontier with the global Malmquist index, which decomposes productivity change into efficiency change, best-practice change within the group and change in the technology gap. To the author's knowledge, these tools have not been applied to national aviation systems.

Inference in two-stage and productivity analysis

Regressions of DEA scores on environmental variables raise inferential problems. Simar and Wilson (2007) showed that conventional inference in such regressions is invalid because the scores are estimated, bounded and serially correlated, and that the second stage has a clear interpretation only under a separability condition, under which the environmental variables affect the distribution of efficiency but not the shape of the frontier. They proposed a truncated regression with a parametric bootstrap and a double bootstrap that also corrects the scores for bias, both developed for radial measures. Other authors argue that simpler estimators, such as fractional response models (Papke and Wooldridge, 1996), can be consistent for the conditional mean of the scores (Banker and Natarajan, 2008; McDonald, 2009); the debate concerns mainly what the second stage estimates and how its inference is carried out (Simar and Wilson, 2011a). For productivity indices, Simar and Wilson (1999) proposed a bootstrap for Malmquist indices. Resampling with replacement is not consistent for frontier estimators, whereas subsampling is (Simar and Wilson, 2011b), and when the number of inputs and outputs is large relative to the sample the bias of DEA estimators can be of the same order as their standard deviation (Kneip et al., 2015). These results shape the inference used here and the caution with which its intervals are read.

Methods

Research design

The study uses a quantitative, non-experimental, longitudinal design based on secondary country-level panel data. National aviation systems are the units of analysis, observed annually over 2010–2023. The analysis has three steps: environmental efficiency is measured with a slacks-based DEA model against a metafrontier and income-group frontiers; productivity change is measured and decomposed with a metafrontier global Malmquist–Luenberger index; and the efficiency scores are related to country conditions in a second-stage regression. Inference for all three steps uses a country-block bootstrap. The design is descriptive and associational: it benchmarks performance and identifies correlates of efficiency but does not estimate causal effects.

Sample

The target population is all economies with commercial air transport. Of the 217 economies in the assembled panel, 155 have strictly positive values of all DEA variables in at least one year. Because the SBM scores of microstates and of systems with negligible commercial aviation are dominated by near-zero outputs, the main sample is restricted, by a criterion fixed before estimation, to economies with an average population of at least one million and an average of at least one million passengers a year. This criterion-based selection leaves 84 countries and 1,156 country-years in an unbalanced panel; the number of years per country ranges from 5 to 14 and is 14 for most countries. Countries are grouped by the current World Bank income classification into high-income (40 countries), upper-middle-income (25) and lower-middle- and low-income countries (19). The two low-income countries, Afghanistan and Ethiopia, are merged with the lower-middle-income group because a frontier estimated from two countries would be uninformative. The 155-economy sample, with 54, 43 and 58 economies in the three groups, is used as a robustness check.

Data collection

All data are secondary and publicly available; no primary data were collected. Traffic, macroeconomic and emissions series were taken from the World Development Indicators (WDI; World Bank, 2026), the jet fuel spot price from the U.S. Energy Information Administration (U.S. EIA, 2026) and airport records from the OurAirports database (OurAirports, 2026), all accessed in September–October 2026, and merged by ISO 3166-1 alpha-3 country code and year. Departures, passengers and freight are WDI series compiled

from International Civil Aviation Organization (ICAO) statistics and refer to the worldwide operations of carriers registered in each country, so countries that host the base of a large international carrier record traffic far above their domestic market. No public emissions series has the same boundary. Country-level aviation emissions in the Emissions Database for Global Atmospheric Research (EDGAR) cover only domestic flights; in the release examined (European Commission JRC, 2026), international aviation is a single unallocated item. The main specification therefore uses transport-sector CO_2, as in Queiroz Júnior et al. (2024).

Variables and measures

The DEA model has two inputs, two desirable outputs and one undesirable output (Table 1). The inputs are the number of large and medium airports with scheduled commercial service in OurAirports, a time-invariant snapshot, and departures of carriers registered in the country. The desirable outputs are passengers carried and air freight in million tonne-kilometres. The undesirable output is CO_2 from the transport sector (WDI series EN.GHG.CO2.TR.MT.CE.AR5, based on EDGAR); because its boundary is wider than aviation, the scores are environmental efficiency scores conditional on the national transport-carbon context. The second-stage covariates are GDP per capita in constant 2015 US dollars, population and energy use per capita, all in logarithms, CO_2 per capita, a real jet fuel price, a COVID-19 dummy for 2020–2021 and a linear trend. The real fuel price is the U.S. Gulf Coast jet fuel spot price converted at the official exchange rate and deflated by the consumer price index; because the world price varies only over time, its level is not comparable across countries, and it enters as the within-country deviation of its logarithm. Observations with exchange-rate regime breaks or missing covariates are excluded from the second stage, which has 1,079 observations for 79 countries.

Data analysis

Slacks-based measure with an undesirable output

Consider n country-years, each using m inputs x, producing s_1 desirable outputs y^g and s_2 undesirable outputs y^b. Following Tone (2004), the SBM efficiency of unit o is the solution of equation (1) subject to the constraints in equation (2):

$$E_o^* = \min \frac{1 - \frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{io}}{1 + \frac{1}{s_1 + s_2}\left(\sum_{r=1}^{s_1} s_r^g / y_{ro}^g + \sum_{k=1}^{s_2} s_k^b / y_{ko}^b\right)} \qquad (1)$$

$$x_o = X\lambda + s^-, \quad y_o^g = Y^g\lambda - s^g, \quad y_o^b = Y^b\lambda + s^b, \quad \sum_j \lambda_j = 1, \quad \lambda, s^-, s^g, s^b \geq 0 \qquad (2)$$

where X, Y^g and Y^b are the matrices of inputs, desirable outputs and undesirable outputs of the reference units, λ is the vector of intensity weights and s^-, s^g and s^b are the slacks. The condition that the intensity weights sum to one imposes variable returns to scale (VRS) and is dropped under constant returns to scale (CRS). The score lies in (0, 1]; a score of 1 means that no input can be reduced, no desirable output increased and no emission reduced without worsening another dimension. The fractional programme is solved as a linear programme through the Charnes–Cooper transformation (Cooper et al., 2007). VRS is the main specification because the size of a national aviation market is set largely by population, income and geography; CRS scores are reported as a robustness check. Inputs and outputs are divided by their sample means before estimation, to which SBM scores are invariant.

Metafrontier and technology gap

Three reference sets are used for each country-year. The metafrontier score E^M is computed against all 1,156 observations of all groups and years. The intertemporal group score E^I is computed against all observations of the country's income group, and the contemporaneous group score E^C against the observations of the same group in the same year. Because each reference set contains the next, $E^M \leq E^I \leq E^C$. The technology gap ratio (TGR) of equation (3) measures how close the best practice of a country's group comes to the overall best practice at the country's input–output mix (O'Donnell et al., 2008):

$$TGR_{it} = E_{it}^M / E_{it}^I \in (0,1] \qquad (3)$$

A TGR of 1 means that the group frontier touches the metafrontier at that point; a low TGR means that the best practice available within the group lies far inside the metafrontier. The metafrontier score is the product of two components, efficiency relative to the group frontier and the technology gap. In this context the term technology refers to everything that makes a group's frontier differ from the metafrontier, including network structure, traffic mix and geography, not only aircraft or airport technology.

Productivity change

Combining the global Malmquist–Luenberger (GML) index with undesirable outputs (Oh, 2010) with the metafrontier decomposition of Oh and Lee (2010), productivity change between $t-1$ and t is measured on the metafrontier and decomposed as in equation (4):

$$GML_{t-1,t} = \frac{E_t^M}{E_{t-1}^M} = EC \times BPC \times TGC, \quad EC = \frac{E_t^C}{E_{t-1}^C}, \quad BPC = \frac{E_t^I / E_t^C}{E_{t-1}^I / E_{t-1}^C}, \quad TGC = \frac{E_t^M / E_t^I}{E_{t-1}^M / E_{t-1}^I} \qquad (4)$$

EC measures catch-up to the contemporaneous group frontier, BPC the movement of the group's contemporaneous best practice towards its intertemporal frontier, and TGC the change in the technology gap. Values above 1 indicate improvement. The product BPC × TGC is the frontier-shift component, denoted TC. Indices are computed for consecutive years only (1,068 annual changes), and averages are geometric means. For each period, D = mean ln EC − mean ln TC is positive when catch-up exceeds frontier shift and negative when frontier shift exceeds catch-up.

Second stage

The second stage relates the metafrontier score to country conditions with the fractional logit model of equation (5), estimated by quasi-maximum likelihood (Papke and Wooldridge, 1996), where $G(\cdot)$ is the logistic function and the fuel price term is the within-country deviation of the log real fuel price:

$$E(E_{it}^M \mid z_{it}) = G\big(\beta_0 + \beta_1 \ln GDPpc_{it} + \beta_2 \ln Pop_{it} + \beta_3 \ln Energy_{it} + \beta_4 CO2pc_{it} + \beta_5 \widetilde{FP}_{it} + \beta_6 COVID_t + \beta_7 t\big) \qquad (5)$$

As an alternative that follows the logic of Algorithm 1 of Simar and Wilson (2007), a truncated regression is estimated with $-\ln E^{\mathrm{M}}$, which is zero on the frontier, as the dependent variable and a normal error truncated at zero, on the observations below the frontier. The Farrell-type measure $1/E$ is not used because SBM scores close to zero make it extremely skewed. Positive coefficients in this model indicate lower efficiency.

Statistical inference

Inference for all quantities uses a country-block bootstrap of the whole procedure. In each replicate, countries are drawn with replacement within each income group, so that group sizes are preserved, and each country enters with its full time series. The metafrontier and both group frontiers are re-estimated on the pseudo-sample, the scores, technology gap ratios and productivity components are recomputed, and both second-stage models are re-estimated. The bootstrap distribution reflects the estimation of the frontiers and scores as well as within-country dependence. Intervals are 95% percentile intervals and p-values are twice the smaller

share of replicates on either side of zero (of zero in logarithms for the indices). All tests are two-sided with $\alpha = 0.05$; an estimate is described as significant when its bootstrap p-value is below 0.05, and significance at the 10% level is stated explicitly. Differences in country-mean scores across income groups are also tested with Kruskal–Wallis tests, which do not assume normality and suit bounded, skewed scores, with ε^2 as the effect size, and associations between the technology gap ratio and traffic characteristics are measured with Spearman rank correlations ($n = 84$ countries). The naive bootstrap is not consistent for frontier estimators (Simar and Wilson, 1999, 2011b), the bias of DEA averages is not corrected (Kneip et al., 2015), and the separability condition is not tested, so the intervals are approximate and the second-stage coefficients are read as conditional associations, not causal effects.

Robustness checks

Four alternative specifications re-estimate all scores and indices: (i) the 155-economy sample without the size rule, (ii) the main sample without the five economies that host large internationally based carriers and appear in the reference sets of most other countries (Hong Kong, Qatar, the United Arab Emirates, Ireland and Singapore; Luxembourg is already excluded by the size rule), (iii) CRS instead of VRS for the static metafrontier scores and (iv) passengers as the only desirable output. The overlap with an earlier 37-country sample of G20, ASEAN and South American countries is also examined.

Software

All estimates were computed in Python 3.13 (SciPy 1.18 with the HiGHS linear programming solver, pandas 3.0 and statsmodels 0.15). Parts of the estimation code were written with the assistance of Claude (Anthropic); the author checked all code and outputs. All figures are generated directly from the estimation results with Matplotlib.

Reliability, validity and ethics

Several features of the design support the reliability and validity of the results. The sample criterion was fixed before estimation. The SBM code reproduces to machine precision the scores of an earlier 37-country specification, which had been verified against an independent Stata implementation (maximum difference 1.2×10^{-7}), and all results can be regenerated from the supplied data and code with a fixed random seed. Every frontier is re-estimated in every bootstrap replicate, so the reported intervals reflect the estimation of the frontiers. The main threat to construct validity is the boundary mismatch between the traffic and emissions data; it is addressed by interpreting the scores as conditional on the transport-carbon context,

by examining which slack terms drive the scores and by re-estimating the model without the hub economies, with passengers as the only desirable output, under constant returns to scale and in the 155-economy sample. Group frontiers estimated from equal numbers of countries check that the comparison of group-frontier scores is not driven by group size. The study uses only publicly available country-level aggregates and involves no human participants or personal data, so ethical approval and informed consent were not required.

Limitations

The design has several limitations. The traffic data follow the nationality of carriers, so economies that host international carriers record traffic far above their domestic markets and can dominate the metafrontier; airport-based traffic statistics would avoid this but are not available for all countries and years. The emissions variable covers all transport modes rather than aviation alone. The airport count is a single recent snapshot, and passengers are measured without distance while freight is measured in tonne-kilometres, so the choice of outputs shapes the scores. The bootstrap intervals are approximate: the naive bootstrap is not consistent at the frontier, the bias of DEA averages is not corrected and the separability condition is not tested. The income classification is the current one and is applied to all years, and the panel contains only two post-pandemic years. Because the selection criterion excludes small economies, the findings apply to medium and large aviation systems rather than to microstates.

Results

The results are reported in the order of the three research questions: differences in efficiency and in the technology gap across income groups, productivity change, and the country conditions associated with efficiency. Descriptive statistics are reported first and the robustness checks last.

Descriptive statistics

Table 2 reports descriptive statistics for the main sample. The variables spanned several orders of magnitude: scheduled-service airports ranged from 1 to 505, annual passengers from 0.15 million to 942 million and transport CO_2 from 0.9 to 1,762 Mt. Across country means, the median freight output per departure was 5.8 thousand tonne-kilometres in the high-income group, 1.9 in the upper-middle-income group and 2.0 in the lower-middle- and low-income group, and the median number of passengers per scheduled-service airport was 1.39 million, 0.64 million and 0.30 million.

Efficiency and the technology gap across income groups

The mean metafrontier score was 0.149 and the median 0.057; 79% of the observations scored below 0.2. Thirty-three observations from 13 countries lay on the metafrontier: 24 from eight high-income economies (Bahrain, Hong Kong, Ireland, Latvia, Qatar, Singapore, the United Arab Emirates and the United States), eight from four upper-middle-income countries (Azerbaijan, China, Mauritius and Turkmenistan) and one from Lebanon. The median of country means of passengers per inhabitant was 2.38 for these 13 countries and 0.39 for the others. The five hub economies listed under Robustness checks supplied 14 of the 33 frontier observations. Hong Kong was in the reference set of 71% of the 1,123 observations below the metafrontier and Qatar of 54%, and 96% of these observations had at least one of the five hub economies in their reference set.

Table 3 and Fig. 1 report the scores by income group. Mean metafrontier scores fell with income: 0.211 [0.175, 0.331] in the high-income group, 0.127 [0.076, 0.234] in the upper-middle-income group and 0.048 [0.031, 0.162] in the lower-middle- and low-income group, with a difference of 0.163 [0.067, 0.260] (p = 0.004) between the first and the last. Group-frontier scores did not follow this pattern: they were 0.228 [0.191, 0.348], 0.283 [0.247, 0.510] and 0.292 [0.255, 0.582], and the difference between the high-income and the lowest-income group was −0.064 [−0.337, 0.048] (p = 0.168). Across the 84 country means, Kruskal–Wallis tests rejected equal metafrontier scores across groups ($H(2)$ = 10.32, p = 0.006, ε^2 = 0.12) but not equal group-frontier scores ($H(2)$ = 0.94, p = 0.625, ε^2 = 0.01). With group frontiers estimated from 19 randomly drawn high-income or upper-middle-income countries (40 draws each), the mean group-frontier score was 0.285 (2.5th–97.5th percentile of draws 0.227–0.382) for high-income and 0.327 (0.255–0.457) for upper-middle-income countries, compared with 0.292 for the 19 lower-middle- and low-income countries.

The technology gap ratio averaged 0.939 [0.870, 0.981] in the high-income group, where the group frontier coincided with the metafrontier for 71% of the observations, 0.462 [0.174, 0.601] in the upper-middle-income group and 0.165 [0.061, 0.343] in the lower-middle- and low-income group (Kruskal–Wallis $H(2)$ = 64.04, $p < 0.001$, ε^2 = 0.77 across country means). The high-income group differed significantly from the upper-middle-income group (0.477 [0.301, 0.797], $p < 0.004$) and from the lowest-income group (0.774 [0.592, 0.887], $p < 0.004$); the difference between the upper-middle- and lowest-income groups (0.298 [−0.088, 0.468], p = 0.172) was not significant. Within the upper-middle-income group, the ratio ranged from above 0.8 for Turkmenistan, Libya, Azerbaijan, Mauritius and China to below

0.2 for El Salvador and Viet Nam (Supplementary Table S1). Across country means, the technology gap ratio was positively correlated with freight per departure (Spearman $\rho = 0.36$, $p < 0.001$), passengers per airport ($\rho = 0.53$, $p < 0.001$) and passengers per Mt of transport CO_2 ($\rho = 0.34$, $p = 0.002$).

Among the 1,123 observations below the metafrontier, the freight shortfall was the largest of the three terms in the denominator of equation (1) in 92% of cases, the CO_2 excess in 8% and the passenger shortfall in fewer than 1%; the median share of freight in the sum of these terms was 0.98. Freight was the largest term for 87% of the high-income, 95% of the upper-middle-income and 99% of the lower-middle- and low-income observations.

Productivity change

Table 4 reports the metafrontier GML index and its components, and Fig. 2 shows their annual values. Over the 1,068 annual changes, the geometric mean GML was 1.022 [1.000, 1.043], about 2% a year, but the growth was not significant at the 5% level ($p = 0.056$); GML exceeded 1 in 54% of the country-years. BPC averaged 1.068 [1.016, 1.080] ($p < 0.004$), whereas EC (0.973 [0.960, 1.010], $p = 0.248$) and TGC (0.983 [0.970, 1.008], $p = 0.196$) did not differ significantly from 1. In 2011–2019, BPC was 1.064 [1.007, 1.088] ($p = 0.016$), EC 0.973 [0.950, 1.017] and TGC 0.987 [0.970, 1.007]; the difference D between the mean logarithms of EC and of the frontier-shift component was −0.077 [−0.124, 0.010] and was not significant ($p = 0.100$).

Intervals were wide in the pandemic and recovery years. GML was 1.019 [0.797, 1.335] in 2020 and 0.989 [0.820, 1.243] in 2021 and averaged 1.038 [0.891, 1.171] in 2022–2023, with EC 1.083 [0.855, 1.211] and frontier shift 0.959 [0.839, 1.240]. None of these intervals excluded 1, and the change in D between 2011–2019 and 2022–2023 was not significant ($p = 0.740$). By income group, GML growth was significant only in the upper-middle-income group (1.052 [1.006, 1.102], $p = 0.028$), where BPC was also significant (1.077 [1.010, 1.113], $p = 0.012$; Table 4, lower panel).

Country conditions associated with efficiency

Table 5 reports the second-stage estimates. In the fractional logit, GDP per capita (0.794 [0.110, 1.519], $p = 0.004$), CO_2 per capita (0.086 [0.041, 0.248], $p = 0.004$) and the trend (0.026 [0.003, 0.064], $p = 0.036$) were positively associated with the metafrontier score. The average marginal effect of log GDP per capita was 0.095 [0.015, 0.194], so a 1% higher GDP per capita was associated with a score higher by 0.0009, against a mean score of 0.149.

Population, energy use per capita (p = 0.116), the within-country fuel price and the COVID-19 dummy (0.102 [−0.218, 0.432], p = 0.524) were not significant. In the truncated regression of the negative log metafrontier score, where negative coefficients indicate higher efficiency, GDP per capita (−1.458, $p < 0.004$) and CO_2 per capita (−0.225, $p < 0.004$) had the same signs, larger population was associated with higher efficiency (−0.588, p = 0.004) and higher energy use per capita with lower efficiency (1.586, p = 0.032); the COVID-19 coefficient was not significant (p = 0.592).

Re-estimating the fractional logit with the country means of the four time-varying covariates (Mundlak, 1978) separated between-country from within-country variation. The coefficient on the country mean of log GDP per capita was 0.833 [0.134, 1.631] (p = 0.004) and that on its within-country deviation 0.185 [−2.029, 2.018] (p = 0.832). For CO_2 per capita, the between-country coefficient was 0.092 [0.044, 0.269] (p = 0.004) and the within-country coefficient −0.026 [−0.132, 0.071] (p = 0.520).

Robustness

Table 6 reports the main quantities under the alternative specifications. Without the five hub economies, the mean metafrontier score rose to 0.231 and the technology gap ratios were 0.839, 0.715 and 0.323. In this specification (B = 199), the difference in technology gap ratios between the high-income and the lowest-income group remained significant (0.516 [0.197, 0.744]), the difference between the high- and upper-middle-income groups was not significant (0.124 [−0.193, 0.550]), and the group-frontier scores showed no significant difference between the high- and lowest-income groups (0.061 [−0.177, 0.203]). In the 155-economy sample the technology gap ratios were 0.922, 0.379 and 0.173. Under CRS, the mean metafrontier score was 0.090 and the Spearman correlation with the VRS scores was 0.933. For the 30 countries shared with an earlier 37-country sample of G20, ASEAN and South American countries, the Spearman correlation between the metafrontier scores and the scores against the 37-country frontier was 0.99.

With passengers as the only desirable output (Table 6, last column), the technology gap ratios were 0.940, 0.647 and 0.514. Group-frontier scores were 0.370, 0.486 and 0.489 (Kruskal–Wallis $H(2)$ = 10.75, p = 0.005, ε^2 = 0.13), and metafrontier scores were 0.345, 0.307 and 0.245 ($H(2)$ = 4.60, p = 0.100, ε^2 = 0.06). The Spearman correlation between the metafrontier scores with and without freight was 0.475. This specification was not bootstrapped.

Before 2020, frontier shift exceeded catch-up in every specification (frontier shift between 1.032 and 1.051, EC between 0.954 and 0.994). In 2022–2023, EC exceeded 1 in every specification, and the frontier-shift component ranged from 0.929 to 1.108. The point estimates for GDP per capita and CO_2 per capita were positive in all specifications, whereas the sign of the COVID-19 coefficient varied.

Discussion

The first research question asked whether the lower environmental efficiency of lower-income aviation systems reflects inefficiency relative to comparable systems or a gap in the best practice available to them. The results point to the second. Against a single frontier, lower-income countries appear far less efficient; against the frontier of their own income group, they do not, including when group frontiers are estimated from equal numbers of countries. Their low metafrontier scores therefore reflect the position of their group's best practice, which lies far inside the metafrontier at their input–output mix, rather than a larger distance from comparable systems. In metafrontier terms this is a technology gap, but the model does not identify what the gap consists of. The ratio is higher where freight per departure, passengers per airport and passengers per Mt of transport CO_2 are higher, which points to traffic mix, network concentration and the carbon intensity of the wider transport sector as candidate explanations; the data contain no fleet, load-factor or network variables with which to test them.

Two features of the data qualify this reading. First, most measured inefficiency is a shortfall in freight output relative to frontier units with very large cargo operations; the emissions variable shapes the frontier and the choice of peers but is rarely the main source of inefficiency. Without freight, the gradient in overall scores largely disappears while the technology gap remains ordered by income, so country-level comparisons that use freight tonne-kilometres as an output largely compare cargo operations. Second, because the traffic data follow the nationality of carriers, economies that host large international carriers record very high output per airport (a median of 14.9 million passengers against 0.62 million for the other countries) and dominate the reference sets. Without them, the gap between high- and upper-middle-income countries narrows and is no longer significant, so part of that gap separates hub economies from all others; the gap of the lowest-income group remains.

The comparison with earlier evidence is informative. Queiroz Júnior et al. (2024) found the United States, China, India and Germany fully efficient in all years when GDP and population

were part of the technology. Without these aggregates, the United States and China score far above the sample mean of 0.149 (0.845 and 0.669), whereas Germany (0.254) and India (0.101) do not, which is consistent with the concern that macroeconomic aggregates inside the technology favour large economies. Airline-level studies found sharp falls in efficiency in 2020 (Li et al., 2022; Voltes-Dorta et al., 2024; Wu et al., 2024), whereas neither second-stage model detects a pandemic effect at the national level and the productivity indices for 2020 and 2021 have wide intervals. One possible reason is that all national systems lost traffic at the same time, so the decline shifted the frontier as well as the units measured against it.

On the second research question, point estimates attribute productivity growth to improvements in best practice within the income groups rather than to catch-up, but overall growth is not significant at the 5% level, and the pandemic and recovery years are too few and too variable for the decomposition to be informative. On the third, GDP per capita and CO_2 per capita are associated with efficiency across countries but not within countries over time, so the associations reflect differences between countries rather than changes within them. GDP per capita is closely related to the income grouping that defines the technology gap, so its between-country coefficient may largely restate the gradient in metafrontier scores. The positive coefficient on CO_2 per capita does not mean that higher emissions raise efficiency: it holds conditional on GDP per capita and energy use. In a cross-country regression of mean CO_2 per capita on log GDP per capita and log energy use, the largest positive residuals belong to Qatar, Bangladesh, Kuwait, Nigeria, the United Arab Emirates, Australia and Bahrain; three of these lie on the metafrontier, whereas Bangladesh and Nigeria have low scores, so the association has no single obvious source.

For policy, the results caution against reading single-frontier rankings of national aviation systems as evidence of poor management in lower-income countries. Within their own income groups, lower-income systems are on average as close to best practice as high-income ones. The gap that matters for these countries is the distance between their group’s best practice and the global one, which may depend on network structure, traffic mix and access to long-haul and cargo markets rather than on the decisions of individual operators. Efficiency comparisons used in climate or infrastructure policy would be more informative if they reported within-group performance and the technology gap separately.

These interpretations are subject to the limitations set out in the Methods. In particular, the traffic data are carrier-based while the emissions data cover the whole transport sector, so the

scores measure efficiency conditional on the national transport-carbon context rather than aviation-specific carbon efficiency.

Conclusions and further research

Rankings of national aviation systems can shape views about which countries need to improve their environmental performance, yet a ranking against a single global frontier cannot distinguish a poorly run system from one that operates under different structural conditions. This study separated the two for 84 countries over 2010–2023, using an income-group metafrontier, a decomposition of productivity change and bootstrap inference.

Lower-income aviation systems are not less efficient than comparable systems; their low scores against the global frontier reflect a gap between the best practice available to their income group and the global one. For the lowest-income group, this gap persists when the hub economies that dominate the frontier are excluded, although most measured inefficiency reflects freight output rather than emissions. Point estimates attribute productivity growth to shifts in best practice rather than to catch-up, but overall growth is not significant at the 5% level. Income and per-capita emissions are associated with efficiency across countries but not within countries over time.

For governments and international bodies that benchmark national aviation systems, the main implication is to report within-group efficiency and the technology gap separately and to treat freight-dominated scores with caution when environmental performance is the concern. Narrowing the gap for lower-income countries may depend more on network structure, traffic mix and market access than on the decisions of individual carriers. Future research could explain the technology gap with fleet, load-factor and network variables; replace carrier-based traffic and transport-sector emissions with airport-based traffic and carrier-attributed emissions, for example from flight-level automatic dependent surveillance–broadcast (ADS-B) inventories (Wang et al., 2024), which would remove both the hub effect and the boundary mismatch; extend the panel to more post-pandemic years; and develop bias-corrected inference for metafrontier technology gap ratios.

References

Arjomandi A, Seufert JH (2014) An evaluation of the world's major airlines' technical and environmental performance. Econ Model 41:133–144.
https://doi.org/10.1016/j.econmod.2014.05.002

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30:1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Banker RD, Natarajan R (2008) Evaluating contextual variables affecting productivity using data envelopment analysis. Oper Res 56:48–58. https://doi.org/10.1287/opre.1070.0460

Battese GE, Rao DSP, O'Donnell CJ (2004) A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. J Product Anal 21:91–103. https://doi.org/10.1023/B:PROD.0000012454.06094.29

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Chung YH, Färe R, Grosskopf S (1997) Productivity and undesirable outputs: a directional distance function approach. J Environ Manag 51:229–240. https://doi.org/10.1006/jema.1997.0146

Cooper WW, Seiford LM, Tone K (2007) Data envelopment analysis: a comprehensive text with models, applications, references and DEA-Solver software, 2nd edn. Springer, New York. https://doi.org/10.1007/978-0-387-45283-8

European Commission JRC (Joint Research Centre) (2026) IEA-EDGAR CO_2 emissions by country and sector, 1970–2025. Emissions Database for Global Atmospheric Research. https://edgar.jrc.ec.europa.eu/dataset_ghg2026. Accessed Oct 2026

Huang F, Zhou D, Hu J-L, Wang Q (2020) Integrated airline productivity performance evaluation with CO_2 emissions and flight delays. J Air Transp Manag 84:101770. https://doi.org/10.1016/j.jairtraman.2020.101770

Kaya Aydin G, Aydin U (2022) Is there any convergence in the CO_2 emission efficiency of airlines? Environ Sci Pollut Res 29:17811–17820. https://doi.org/10.1007/s11356-021-17121-1

Kneip A, Simar L, Wilson PW (2015) When bias kills the variance: central limit theorems for DEA and FDH efficiency scores. Econom Theory 31:394–422. https://doi.org/10.1017/S0266466614000413

Li Y, Huang X-C, Cui Q (2022) Exploring the effect of COVID-19 on airline environmental efficiency through an interval epsilon-based measure model. Environ Sci Pollut Res 29:25623–25638. https://doi.org/10.1007/s11356-021-17610-3

McDonald J (2009) Using least squares and tobit in second stage DEA efficiency analyses. Eur J Oper Res 197:792–798. https://doi.org/10.1016/j.ejor.2008.07.039

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

Ngo T-T, Huynh NT, Yang C-C (2026) A sequential approach to airport efficiency evaluation: integrating two-stage DEA and truncated regression. J Air Transp Manag 131:102922. https://doi.org/10.1016/j.jairtraman.2025.102922

O'Donnell CJ, Rao DSP, Battese GE (2008) Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. Empir Econ 34:231–255. https://doi.org/10.1007/s00181-007-0119-4

Oh D-H (2010) A global Malmquist–Luenberger productivity index. J Product Anal 34:183–197. https://doi.org/10.1007/s11123-010-0178-y

Oh D-H, Lee J-D (2010) A metafrontier approach for measuring Malmquist productivity index. Empir Econ 38:47–64. https://doi.org/10.1007/s00181-009-0255-0

OurAirports (2026) OurAirports open data: airports.csv, countries.csv. https://ourairports.com/data/. Accessed Sept 2026

Papke LE, Wooldridge JM (1996) Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J Appl Econom 11:619–632. https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1

Pastor JT, Lovell CAK (2005) A global Malmquist productivity index. Econ Lett 88:266–271. https://doi.org/10.1016/j.econlet.2005.02.013

Queiroz Júnior HS, Celestino MAS, Souza CFF, Falcão VA, Camioto FC (2024) CO_2 emissions in air transport: a comparative study using data envelopment analysis. Transp Res Rec 2678:872–883. https://doi.org/10.1177/03611981231193407

Simar L, Wilson PW (1999) Estimating and bootstrapping Malmquist indices. Eur J Oper Res 115:459–471. https://doi.org/10.1016/S0377-2217(97)00450-5

Simar L, Wilson PW (2007) Estimation and inference in two-stage, semi-parametric models of production processes. J Econom 136:31–64. https://doi.org/10.1016/j.jeconom.2005.07.009

Simar L, Wilson PW (2011a) Two-stage DEA: caveat emptor. J Product Anal 36:205–218. https://doi.org/10.1007/s11123-011-0230-6

Simar L, Wilson PW (2011b) Inference by the m out of n bootstrap in nonparametric frontier models. J Product Anal 36:33–53. https://doi.org/10.1007/s11123-010-0200-4

Tone K (2001) A slacks-based measure of efficiency in data envelopment analysis. Eur J Oper Res 130:498–509. https://doi.org/10.1016/S0377-2217(99)00407-5

Tone K (2004) Dealing with undesirable outputs in DEA: a slacks-based measure (SBM) approach. Paper presented at the North American Productivity Workshop III, Toronto, 2004

U.S. EIA (U.S. Energy Information Administration) (2026) Spot prices for crude oil and petroleum products: U.S. Gulf Coast kerosene-type jet fuel spot price FOB, annual. https://www.eia.gov/dnav/pet/pet_pri_spt_s1_a.htm. Accessed Sept 2026

Voltes-Dorta A, Britto R, Wilson B (2024) Efficiency of global airlines incorporating sustainability objectives: a Malmquist-DEA approach. J Air Transp Manag 119:102634. https://doi.org/10.1016/j.jairtraman.2024.102634

Wang J, Wang Y, Zhang S, Fan C, Zhou N, Liu J, Li X, Liu Y, Hou X, Yi B (2024) Accounting of aviation carbon emission in developing countries based on flight-level ADS-B data. Appl Energy 358:122600. https://doi.org/10.1016/j.apenergy.2023.122600

World Bank (2026) World development indicators. https://databank.worldbank.org/source/world-development-indicators. Accessed Sept 2026

Wu S, Kremantzis MD, Tanveer U, Ishaq S, O'Dea X, Jin H (2024) Performance evaluation of the global airline industry under the impact of the COVID-19 pandemic: a dynamic network data envelopment analysis approach. J Air Transp Manag 118:102597. https://doi.org/10.1016/j.jairtraman.2024.102597

Yakath Ali NS, Yu C, See KF (2021) Four decades of airline productivity and efficiency studies: a review and bibliometric analysis. J Air Transp Manag 96:102099. https://doi.org/10.1016/j.jairtraman.2021.102099

Figure legends

Figure 1. Metafrontier efficiency, group-frontier efficiency and technology gap ratio by income group. Panel a shows the metafrontier score E^{M}, panel b the intertemporal group-frontier score E^{I} and panel c the technology gap ratio. Each dot is a country mean over 2010–2023; vertical bars mark group medians. Data visualisation generated in Python (Matplotlib) from the estimation results.

Figure 2. Annual metafrontier global Malmquist–Luenberger index and its components, 2011–2023. Panels show the GML index, the catch-up component (EC) and the frontier-shift component (TC = BPC × TGC), each as the geometric mean across countries. Shaded bands are 95% percentile intervals from the country-block bootstrap (B = 500); the dashed line at 1 separates improvement (above) from deterioration (below). Data visualisation generated in Python (Matplotlib) from the estimation results.

Tables

Table 1. Variables and sources

Variable	Definition	Unit	Role	Source
Airports	Large and medium airports with scheduled service	Count	Input	OurAirports
Departures	Departures of registered carriers, worldwide	Count	Input	WDI (IS.AIR.DPRT)
Passengers	Passengers carried by registered carriers	Persons	Desirable output	WDI (IS.AIR.PSGR)
Freight	Air freight	Million tonne-km	Desirable output	WDI (IS.AIR.GOOD.MT.K1)
Transport CO_2	CO_2 from transport (energy), all modes	Mt CO_2e	Undesirable output	WDI/EDGAR (EN.GHG.CO2.TR.MT.CE.AR5)
GDP per capita	GDP / population, constant 2015 US$	US$ (log)	Second stage	WDI (NY.GDP.MKTP.KD, SP.POP.TOTL)
Population	Total population	Persons (log)	Second stage	WDI (SP.POP.TOTL)
Energy use	Energy use per capita	kg oil eq. (log)	Second stage	WDI (EG.USE.PCAP.KG.OE)
CO_2 per capita	CO_2 excluding LULUCF per capita	t CO_2e	Second stage	WDI (EN.GHG.CO2.PC.CE.AR5)
Real fuel price	Jet fuel price × exchange rate / CPI, within-country deviation of the log	Log deviation	Second stage	EIA; WDI (PA.NUS.FCRF, FP.CPI.TOTL)
COVID-19	1 in 2020–2021	Dummy	Second stage	Constructed
Income group	Current World Bank classification (3 groups)	Category	Metafrontier groups	World Bank

Note: WDI = World Development Indicators; EDGAR = Emissions Database for Global Atmospheric Research; EIA = U.S. Energy Information Administration; CPI = consumer price index; LULUCF = land use, land-use change and forestry.

(Source: Author's calculation)

Table 2. Descriptive statistics, main sample (84 countries, 2010–2023)

Variable	N	Mean	SD	Min	Max
Airports (scheduled service)	1156	32.4	63.6	1.0	505.0
Departures (thousand)	1156	373.4	1,104.0	1.4	10,099.0
Passengers (million)	1156	39.21	100.79	0.15	941.56
Freight (million tonne-km)	1156	2,280.814	5,619.360	0.004	46,004.646
Transport CO_2 (Mt)	1156	74.11	205.60	0.92	1,762.24
GDP per capita (constant 2015 US$)	1156	19,042	19,895	378	97,794
Population (million)	1156	78.94	212.32	1.20	1,438.07
Energy use (kg oil eq. per capita)	1128	2,960	2,988	196	21,456
CO_2 per capita (t)	1156	6.90	7.51	0.08	53.60

Note: SD = standard deviation.

(Source: Author's calculation)

Table 3. Metafrontier efficiency, group-frontier efficiency and technology gap ratio by income group

Income group	Metafrontier score E^M	Group-frontier score E^I	Technology gap ratio
High income (40 countries)	0.211	0.228	0.939
	[0.175, 0.331]	[0.191, 0.348]	[0.870, 0.981]
Upper middle income (25)	0.127	0.283	0.462
	[0.076, 0.234]	[0.247, 0.510]	[0.174, 0.601]
Lower middle and low income (19)	0.048	0.292	0.165
	[0.031, 0.162]	[0.255, 0.582]	[0.061, 0.343]
Difference, high minus lowest group	0.163***	−0.064	0.774***
	[0.067, 0.260]	[−0.337, 0.048]	[0.592, 0.887]
Difference, high minus upper middle			0.477***
			[0.301, 0.797]
Difference, upper middle minus lowest			0.298
			[−0.088, 0.468]

Note: means over country-years; brackets: 95% percentile intervals from the country-block bootstrap stratified by income group, with all frontiers re-estimated in every replicate (B = 500). *, ** and *** denote bootstrap $p <$ 0.10, 0.05 and 0.01.

(Source: Author's calculation)

Table 4. Metafrontier global Malmquist–Luenberger index and its components (geometric means)

Period / group	GML	EC	BPC	TGC	TC = BPC × TGC
2011–2019	1.022	0.973	1.064	0.987	1.051
	[0.995, 1.046]	[0.950, 1.017]	[1.007, 1.088]	[0.970, 1.007]	[1.000, 1.075]
2020	1.019	0.945	1.008	1.069	1.078
	[0.797, 1.335]	[0.729, 1.314]	[0.694, 1.350]	[0.783, 1.368]	[0.730, 1.497]
2021	0.989	0.813	1.228	0.991	1.217
	[0.820, 1.243]	[0.701, 1.265]	[0.735, 1.428]	[0.842, 1.313]	[0.761, 1.513]
2022–2023	1.038	1.083	1.038	0.924	0.959
	[0.891, 1.171]	[0.855, 1.211]	[0.906, 1.290]	[0.830, 1.076]	[0.839, 1.240]
2011–2023	1.022	0.973	1.068	0.983	1.050
	[1.000, 1.043]	[0.960, 1.010]	[1.016, 1.080]	[0.970, 1.008]	[1.004, 1.069]
High income, 2011–2023	1.018	0.986	1.047	0.985	
	[0.987, 1.047]	[0.953, 1.028]	[0.995, 1.086]	[0.961, 1.003]	
Upper middle income, 2011–2023	1.052	0.993	1.077	0.983	
	[1.006, 1.102]	[0.954, 1.028]	[1.010, 1.113]	[0.963, 1.033]	
Lower middle and low income, 2011–2023	0.992	0.920	1.100	0.980	
	[0.957, 1.037]	[0.897, 1.022]	[0.990, 1.114]	[0.949, 1.054]	

Note: GML = global Malmquist–Luenberger index on the metafrontier; EC = efficiency change relative to the contemporaneous group frontier; BPC = best-practice change of the group frontier; TGC = technology gap change. Brackets: 95% percentile intervals (B = 500).

(Source: Author's calculation)

Table 5. Second-stage estimates (dependent variable: metafrontier score)

Variable	Fractional logit: coefficient	Fractional logit: average marginal effect	Truncated regression of $-\ln E^{M}$
ln GDP per capita	0.794***	0.0948***	−1.458***
	[0.110, 1.519]	[0.0150, 0.1938]	[−2.375, −0.721]
ln Population	0.090	0.0107	−0.588***
	[−0.395, 0.354]	[−0.0497, 0.0440]	[−0.992, −0.215]
ln Energy use per capita	−0.711	−0.0848	1.586**
	[−2.064, 0.180]	[−0.2725, 0.0232]	[0.191, 3.175]
CO_2 per capita	0.086***	0.0103***	−0.225***
	[0.041, 0.248]	[0.0058, 0.0324]	[−0.404, −0.098]
Real fuel price (within)	−0.033	−0.0040	−0.205
	[−0.283, 0.114]	[−0.0377, 0.0157]	[−0.559, 0.109]
COVID-19 (2020–2021)	0.102	0.0122	0.037
	[−0.218, 0.432]	[−0.0293, 0.0548]	[−0.161, 0.413]
Year trend	0.026**	0.0031**	−0.012
	[0.003, 0.064]	[0.0003, 0.0082]	[−0.043, 0.025]
Observations / countries	1079 / 79	1079 / 79	1047 / 79

Note: brackets: 95% percentile intervals from the country-block bootstrap of the whole two-stage procedure (B = 500). *, ** and *** denote bootstrap $p < 0.10$, 0.05 and 0.01. In the truncated regression, positive coefficients indicate lower efficiency, and the trend is centred on 2016.

(Source: Author's calculation)

Table 6. Robustness of the main results (point estimates)

Quantity	Main	Without hub economies	155 economies	Passengers only
Countries / observations	84 / 1,156	79 / 1,086	155 / 1,887	84 / 1,156
Mean metafrontier score	0.149	0.231	0.097	0.311
TGR, high income	0.939	0.839	0.922	0.940
TGR, upper middle income	0.462	0.715	0.379	0.647
TGR, lower middle and low income	0.165	0.323	0.173	0.514
Group-frontier score, high income	0.228	0.352	0.185	0.370
Group-frontier score, lower middle and low income	0.292	0.292	0.182	0.489
EC, 2011–2019	0.973	0.988	0.954	0.994
TC, 2011–2019	1.051	1.046	1.039	1.032
EC, 2022–2023	1.083	1.081	1.193	1.022
TC, 2022–2023	0.959	1.080	0.929	1.108
Fractional logit: ln GDP per capita	0.794	0.817	1.033	0.196
Fractional logit: CO_2 per capita	0.086	0.117	0.095	0.038
Fractional logit: COVID-19	0.102	−0.242	−0.158	−0.320

Note: TGR = technology gap ratio; EC = efficiency change; TC = frontier shift (BPC × TGC). Hub economies: Hong Kong, Qatar, the United Arab Emirates, Ireland and Singapore. In the 155-economy sample, the lowest group contains 58 lower-middle- and low-income economies. Passengers only: freight removed from the desirable outputs.

(Source: Author's calculation)

Data availability

The raw data analysed during the current study are openly available from the World Bank World Development Indicators (https://databank.worldbank.org/source/world-development-indicators), the U.S. Energy Information Administration (https://www.eia.gov/dnav/pet/pet_pri_spt_s1_a.htm) and OurAirports (https://ourairports.com/data/). The datasets generated during the current study (the assembled panel) are provided as Supplementary Information for peer review and will be deposited in a public repository with a persistent identifier upon acceptance.

Code availability: the Python code used for all estimates, bootstrap procedures, tables and figures is provided as Supplementary Information for peer review and will be released in a public repository with a DOI upon acceptance, without restrictions on access.

Author contributions

The sole author conceived and designed the study, assembled the data, carried out the analysis, interpreted the results and wrote the manuscript.

Competing interests

The author declares no competing interests.

Ethical approval

Ethical approval was not required as no human participants or their data were involved in this research. The study uses only publicly available country-level statistics.

Informed consent

Not applicable, as the study did not involve human participants.

Supplementary information

Supplementary Table S1 (mean scores and productivity change by country, 2010–2023) is provided as a separate Supplementary Information file. The data and code files described under Data availability are supplied as separate supplementary data files.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Unconfirmed254854.crdownload