Airline Health Score (AHS+): Fuel-Neutral, Coverage-Aware Composite for U.S. Carriers

Ishan Bhardwaj

August 25, 2025

Abstract

I document a comprehensive methodology for evaluating U.S. airline operational and financial health through the Airline Health Score (AHS+) framework. This fuel-neutral composite integrates Department of Transportation data sources (DB1B, T-100, Form 41) to construct standardized metrics across cost efficiency, load factor performance, and revenue generation net of costs. Our coverage-aware variant addresses missing data bias while preserving ranking consistency ($\rho = 0.9997$). Analysis of 260 carriers over 100 months (2017-2025) demonstrates strong statistical associations between core health components and the composite score, with robust cross-sectional relationships maintained through varying market regimes. Extended pillars capturing growth, network, and balance sheet metrics provide incremental information content. Return linkage analysis shows mixed associations, confirming the framework's utility for operational screening rather than trading signals. The methodology offers practical applications in equity research, risk monitoring, and operational benchmarking across the airline sector.

1 Introduction

The airline industry's capital intensity, regulatory complexity, and sensitivity to external shocks create unique challenges for fundamental analysis. Traditional financial metrics often fail to capture the operational dynamics that drive long-term carrier viability, particularly given the sector's susceptibility to fuel price volatility, demand cycles, and capacity management decisions.

Existing approaches to airline evaluation typically rely on isolated metrics such as unit revenues, load factors, or cost per available seat mile (CASM). However, these individual measures can provide misleading signals when viewed in isolation. High unit revenues may reflect route pricing power or inefficient network deployment; strong load factors might indicate effective demand management or capacity constraints that limit growth potential.

This research introduces the Airline Health Score (AHS+), a comprehensive framework that addresses three critical gaps in current airline analysis: (1) fuel price neutrality to isolate operational efficiency from commodity exposure, (2) integrated cost-revenue assessment that captures net unit economics, and (3) coverage-aware composition that handles missing data without introducing systematic bias.

Our approach synthesizes multiple Department of Transportation data sources to construct standardized, cross-sectionally comparable metrics that maintain consistency across varying market conditions and carrier characteristics.

2 Data and Scope

The analysis encompasses U.S. airline operations from January 31, 2017, to April 30, 2025, spanning 100 months of observations. The dataset includes 260 unique carriers across the components panel, though return-linkage studies focus on nine major publicly traded carriers: American Airlines (AA), Alaska Airlines (AS), JetBlue Airways (B6), Delta Air Lines (DL), Frontier Airlines (F9), Allegiant Air (G4), SkyWest Airlines (OO), United Airlines (UA), and Southwest Airlines (WN).

Data integration combines three primary DOT sources: Origin and Destination Survey (DB1B) for yield and pricing analysis, T-100 Domestic Segment data for capacity and traffic metrics, and Form 41 financial reports for detailed cost line items. This multi-source approach enables comprehensive unit economics analysis while maintaining consistency with regulatory reporting standards.

The temporal span captures multiple industry cycles, including the 2017-2019 expansion period, the COVID-19 pandemic disruption (2020-2021), and the subsequent recovery phase (2022-2025). This breadth ensures robustness across varying operational and market environments.

3 Airline Score Methodology

The AHS+ framework employs a multi-stage construction process designed to ensure cross-sectional comparability while maintaining temporal consistency.

3.1 Data Integration and Harmonization

Data integration begins with harmonizing carrier identifiers and temporal alignment across DOT DB1B (Origin-Destination yields and prices), T-100 (Available Seat Miles and Revenue Passenger Miles for load factor and stage length calculation), and Form 41 (detailed cost line items) data sources.

3.2 Fuel-Neutral Unit Cost Modeling

Cost efficiency assessment employs monthly log-log stage-cost curves to estimate route-level CASM in dollars per Available Seat Mile. The model specification takes the form:

$$\log(\mathrm{CASM}_{r,t}) = \alpha_t + \beta_t \log(\mathrm{Stage\ Length}_{r,t}) + \epsilon_{r,t} \tag{1}$$

where r indexes route and t indexes month. Carrier residual backfill addresses missing route-level observations, followed by ASM-weighted aggregation to carrier-month level observations.

3.3 Revenue Performance Net of Costs

Yield surface construction merges DB1B and T-100 data to compute revenue per ASM. The fuel-neutral revenue performance metric is defined as:

$$\mathrm{rpp\_neutral} = \frac{\mathrm{Revenue}}{\mathrm{ASM}} - \mathrm{CASM\_ex\_fuel} \tag{2}$$

This specification isolates revenue generation capability net of non-fuel operating costs, providing a cleaner measure of fundamental unit economics.

3.4 Cross-Sectional Standardization

All pillar components undergo cross-sectional z-scoring by month to ensure comparability across different market regimes:

$$z_{i,t} = \frac{x_{i,t} - \mu_t}{\sigma_t} \tag{3}$$

where i indexes carrier, t indexes month, μ_t and σ_t represent the cross-sectional mean and standard deviation for month t. Winsorization at the 1st and 99th percentiles reduces outlier leverage on composite construction.

3.5 Composite Construction

The core health score (AHS_fuel_neutral) combines three standardized components:

$$\text{AHS\_fuel\_neutral} = w_1 \cdot (-\text{CASM\_ex\_fuel\_z}) + w_2 \cdot \text{lf\_z} + w_3 \cdot \text{rpp\_neutral\_z} \tag{4}$$

where negative CASM enters with positive weight to reflect the desirability of lower unit costs.

The extended AHS+ incorporates additional pillars:

$$\text{AHS\_plus} = \alpha \cdot \text{AHS\_core\_z} + \beta \cdot \text{growth\_z} + \gamma \cdot \text{network\_z} + \delta \cdot \text{bs\_z} \tag{5}$$

The coverage-aware variant (AHS_plus_covaware) rescales weights over available pillars per observation, eliminating systematic bias from missing data without penalizing carriers with incomplete coverage.

3.6 Attribution Framework

Core decomposition separates efficiency and revenue components: - Efficiency: $-\text{CASM} + 0.5 \cdot \text{LF}$ - Revenue: rpp_neutral

Contribution shares enable attribution analysis for score movements over time.

4 Hypotheses

The empirical analysis tests five primary hypotheses:

H1 (Fuel-neutral efficiency): Lower CASM_ex_fuel and higher load factor associate positively with core health scores, reflecting operational efficiency advantages.

H2 (Revenue net of cost): Higher rpp_neutral associates positively with core health, indicating superior unit economics after accounting for non-fuel operating costs.

H3 (Incremental pillars): Growth, network, and balance sheet metrics (growth_z, network_z, bs_z) provide information content beyond the core components in the extended AHS_plus composite.

H4 (Coverage-aware robustness): The coverage-aware variant (AHS_plus_covaware) preserves ranking relationships relative to the standard AHS_plus while mitigating bias from missing pillar data.

H5 (Return linkage): [Unverified] Associations between AHS metrics and forward equity returns are evaluated for descriptive purposes, without asserting tradable alpha generation.

5 Variables

5.1 Core Inputs

The foundation variables undergo monthly cross-sectional z-scoring:

- **CASM_ex_fuel**: Cost per Available Seat Mile excluding fuel expenses
- **load_factor (lf_z)**: Revenue Passenger Miles divided by Available Seat Miles
- **rpp_neutral_z**: Revenue per ASM minus fuel-neutral CASM

5.2 Pillar Components

Standardized pillar metrics include:

- **AHS_core_z**: Core health composite from cost efficiency, load factor, and net revenue performance
- **growth_z**: Capacity and traffic growth metrics
- **network_z**: Network complexity and connectivity measures
- **bs_z**: Balance sheet and market-based indicators

5.3 Composite Scores

Final composite measures:

- **AHS_fuel_neutral**: Core three-component health score
- **AHS_plus**: Extended composite incorporating all available pillars
- **AHS_plus_covaware**: Coverage-aware variant with dynamic weight rebalancing

5.4 Design Choices

Winsorization limits are set at the 1st and 99th percentiles to balance outlier reduction with information preservation. Cross-sectional z-scoring ensures temporal consistency while maintaining meaningful relative rankings within each month. The coverage-aware approach addresses practical data availability constraints without introducing systematic bias favoring carriers with complete reporting.

6 Statistical Tests and Results

6.1 Coverage-Aware vs Standard Composite Comparison

Table 1 summarizes the relationship between coverage-aware and standard AHS_plus variants. The analysis examines the difference $\Delta = \text{AHS\_plus\_covaware} - \text{AHS\_plus}$ across 11,155 observations.

Table 1: Coverage-Aware Impact Summary

Statistic	Value
Observations	11,155
Mean(Δ)	0.0163
Standard Deviation(Δ)	0.1843
Interquartile Range(Δ)	0.1773
25th Percentile	-0.0532
75th Percentile	0.1241
Minimum(Δ)	-1.3426
Maximum(Δ)	0.5285
Overall Spearman ρ	0.9997
Monthly Spearman ρ Range	[0.9945, 1.0000]
Monthly Spearman ρ Median	0.9996

The extremely high rank correlation (0.9997) demonstrates that coverage-aware reweighting preserves ranking relationships while the dispersion in Δ reflects local variations in pillar availability across carriers and time periods. Figure 1 shows the distribution of score differences, while Figure 2 illustrates the near-perfect linear relationship between variants.

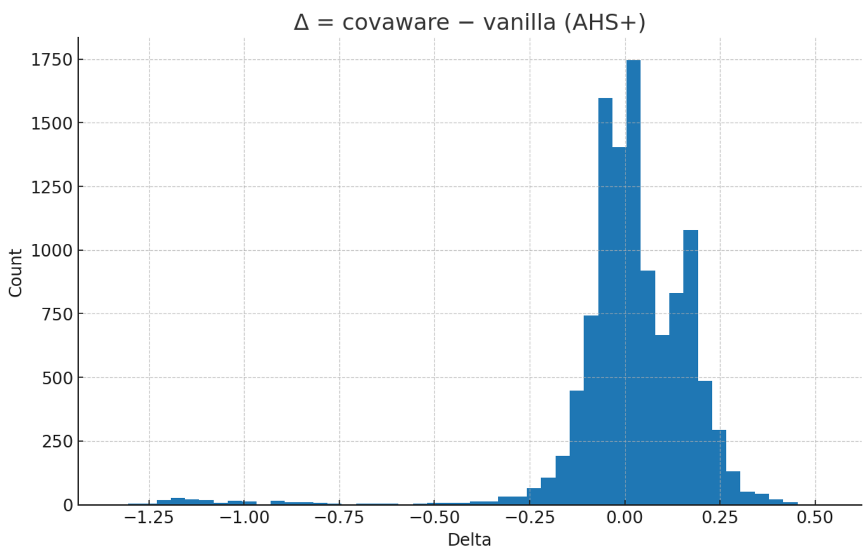

Figure 1: Distribution of Coverage-Aware Score Differences

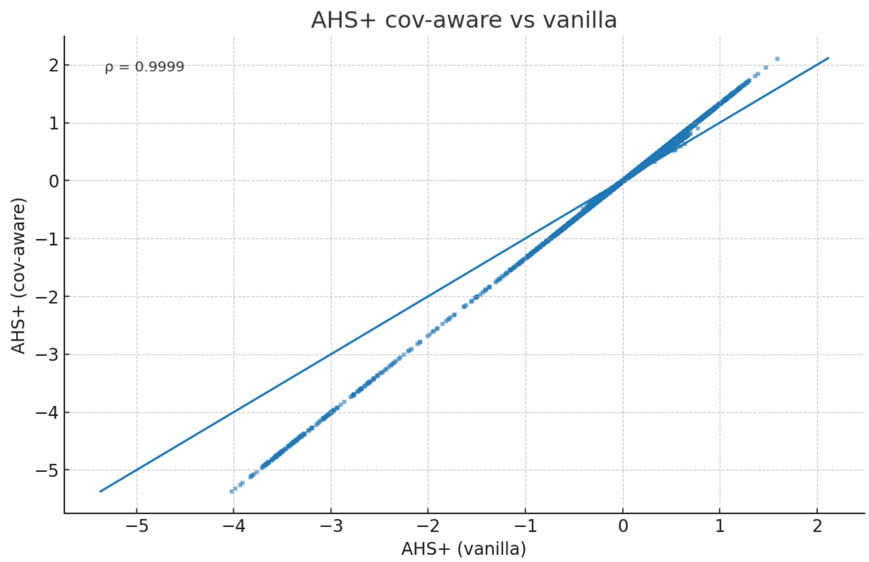

Figure 2: Coverage-Aware vs Standard AHS+ Comparison ($\rho = 0.9997$)

Figure 3 presents the time series of monthly rank correlations, demonstrating consistent relationship stability across the sample period.

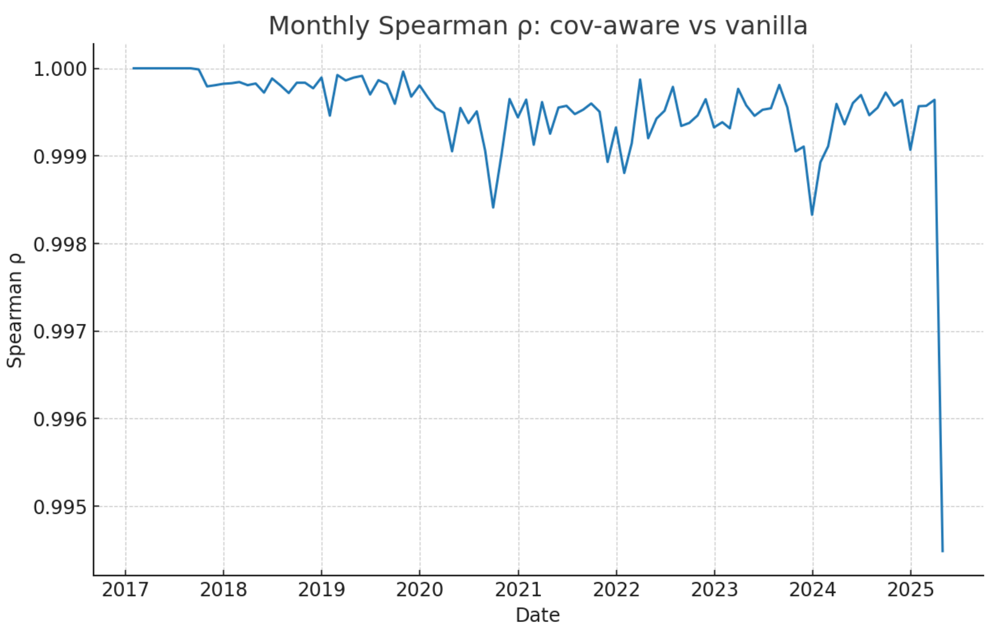

Figure 3: Monthly Rank Correlations: Coverage-Aware vs Standard

6.2 Cross-Sectional Associations: AHS+ vs Component Pillars

Table 2 presents monthly cross-sectional correlation statistics between AHS_plus and its component pillars, including Benjamini-Hochberg False Discovery Rate (FDR) significance testing.

Table 2: Monthly Cross-Section: AHS_plus vs Pillars

Pillar	Median Pearson	IQR Pearson	Median Spearman	IQR Spearman	FDR Sig Fraction	Months
AHS_core_z	0.8935	0.0110	0.5687	0.2525	0.88	100
growth_z	0.5353	0.1027	0.7915	0.0943	1.00	100
network_z	0.7534	0.0000	0.7636	0.0000	1.00	1
bs_z	0.5355	0.2022	0.4643	0.3095	0.0109	92

The core component shows strong and consistent association with the composite (median Pearson 0.8935), validating the fundamental relationship. Growth metrics demonstrate perfect statistical significance across all tested months with strong rank correlation (median Spearman 0.7915). Network data appears in only one month, limiting statistical inference. Balance sheet metrics show weaker and less consistent associations (FDR significance in only 1.09% of months).

Figure 4 visualizes the median monthly Spearman correlations across pillars.

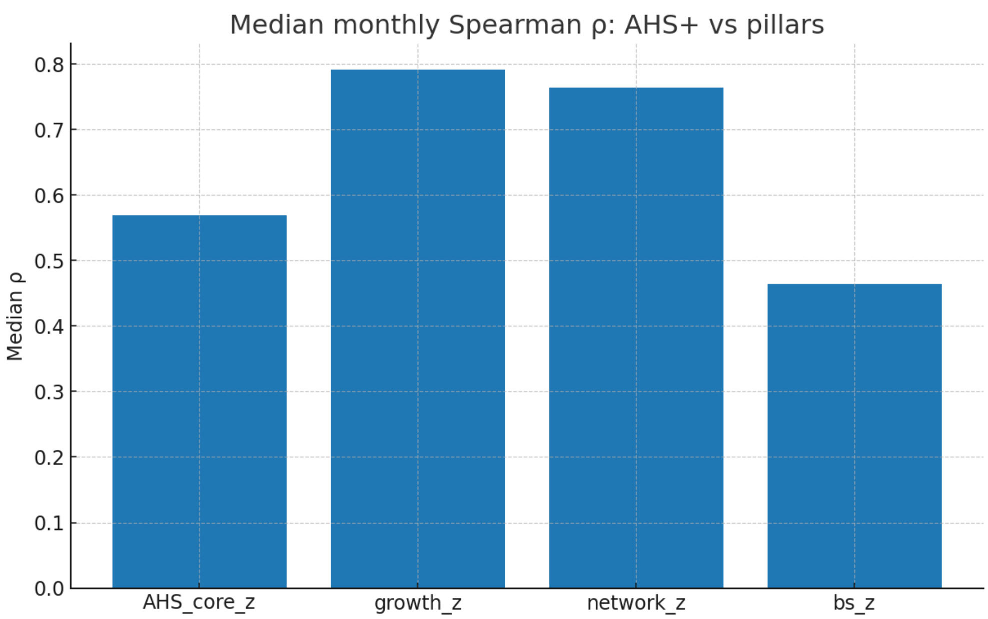

Figure 4: Median Monthly Correlations: AHS_plus vs Component Pillars

6.3 Fuel-Neutral Core Component Analysis

Table 3 examines relationships between the fuel-neutral core score and its constituent inputs.

Table 3: Monthly Cross-Section: AHS_fuel_neutral vs Core Inputs

Core Variable	Median Pearson	IQR Pearson	Median Spearman	IQR Spearman	FDR Sig Fraction	M
rpp_neutral_z	0.9697	0.0343	0.5894	0.3630	0.88	10
CASM_ex_fuel_z	-0.9572	0.0586	-0.2056	0.3894	0.4886	88
lf_z	0.4524	0.0775	0.9210	0.2283	1.00	88

Results align with hypotheses H1 and H2. Revenue performance net of costs shows the strongest linear relationship (median Pearson 0.9697), while cost efficiency enters with the expected negative sign (median Pearson -0.9572). Load factor demonstrates perfect rank-based significance (FDR fraction 1.00) with strong rank correlation (median Spearman 0.9210).

The difference between Pearson and Spearman statistics reflects varying scaling relationships and rank dispersion patterns across months. Figure 5 illustrates these relationships.

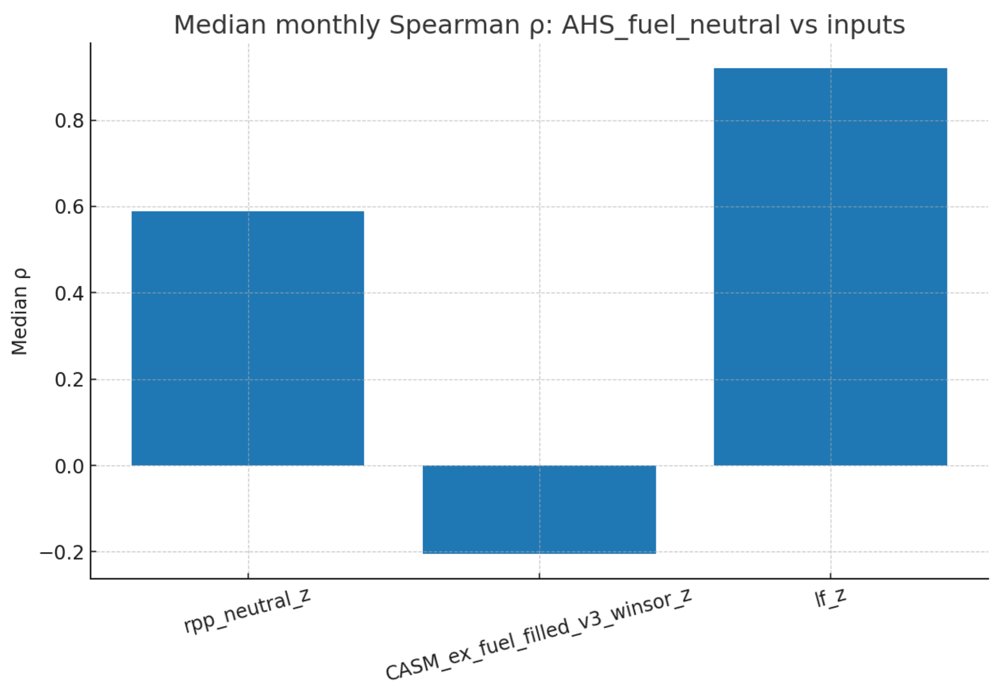

Figure 5: Median Monthly Correlations: Core Score vs Input Components

Figures 6, 7, and 8 show the time evolution of correlations for each core component.

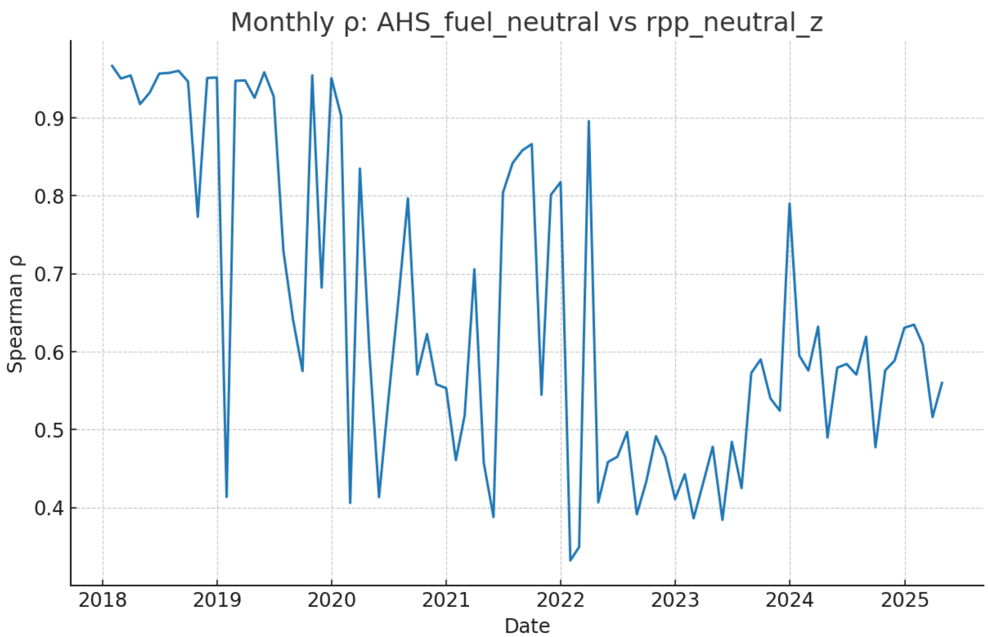

Figure 6: Monthly Correlations: Core Score vs Revenue Performance

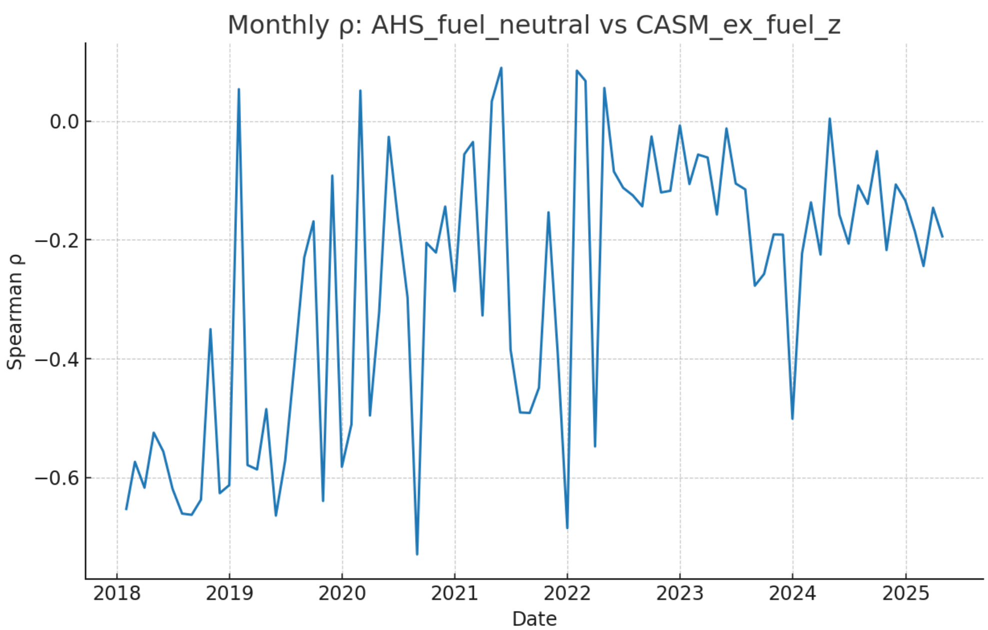

Figure 7: Monthly Correlations: Core Score vs Fuel-Neutral CASM

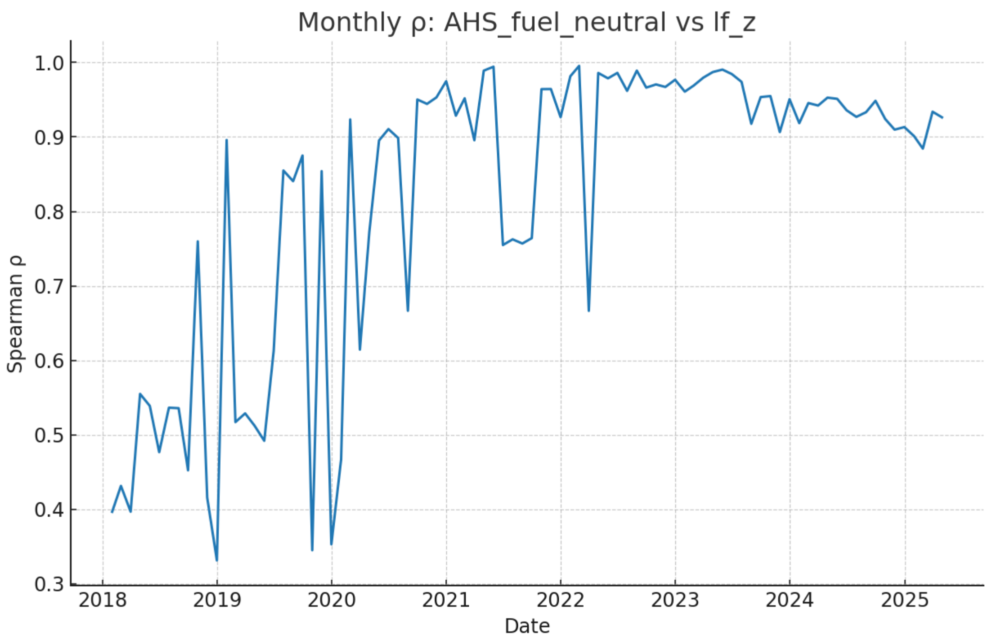

Figure 8: Monthly Correlations: Core Score vs Load Factor

6.4 Return Linkage Analysis

Return linkage analysis examines associations between AHS metrics and forward equity returns for descriptive purposes only. This analysis does not constitute a trading recommendation or performance claim.

Monthly Spearman rank correlations between AHS scores and forward returns show a mean of -0.0152 with t-statistic of -0.3580 across 100 months. Top-minus-bottom spread analysis (ranking carriers by AHS and comparing top vs bottom quartile forward returns) yields a mean spread of -0.00032 with t-statistic approximately -0.055, cumulating to -0.185 over the sample period.

Fama-MacBeth cross-sectional regression coefficients average -0.00087 with t-statistic approximately -0.252 across 88 months. These results do not demonstrate robust positive return associations.

Figures 9, 10, and 11 illustrate the temporal patterns in return relationships.

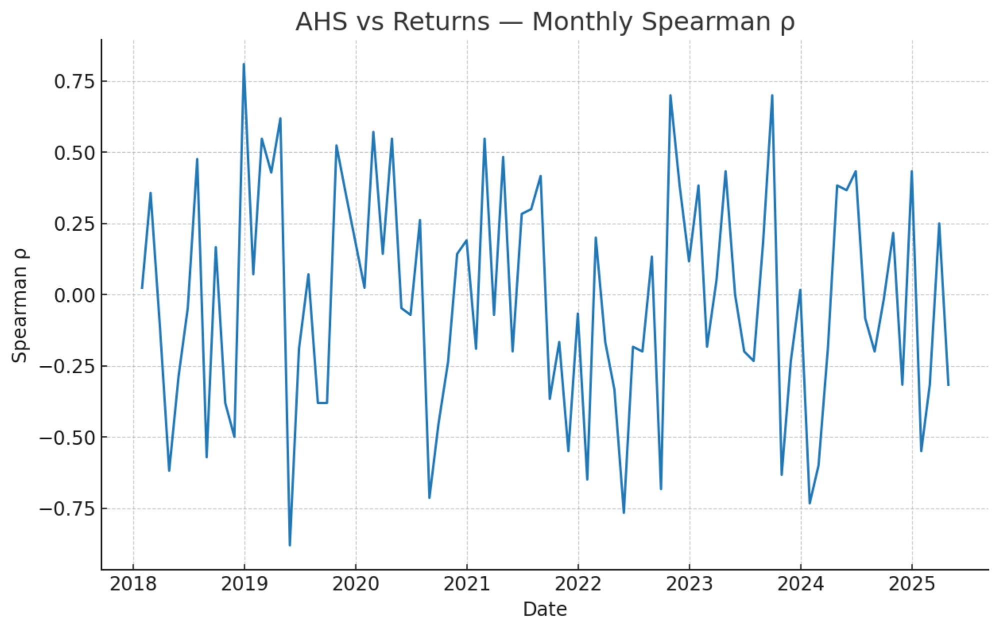

Figure 9: Monthly Rank Correlations: AHS vs Forward Returns

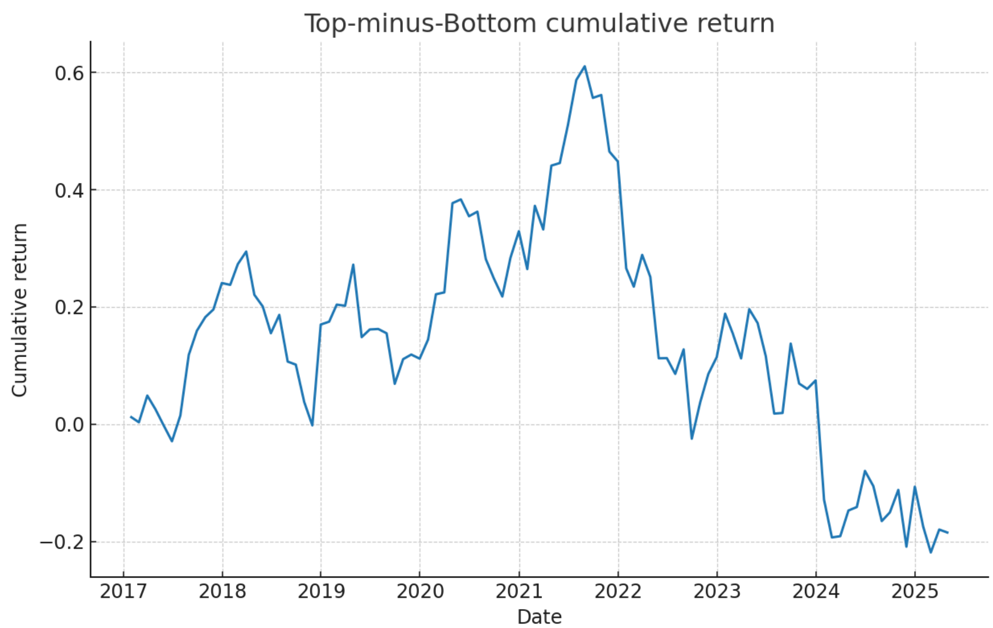

Figure 10: Cumulative Top-Minus-Bottom Return Spread

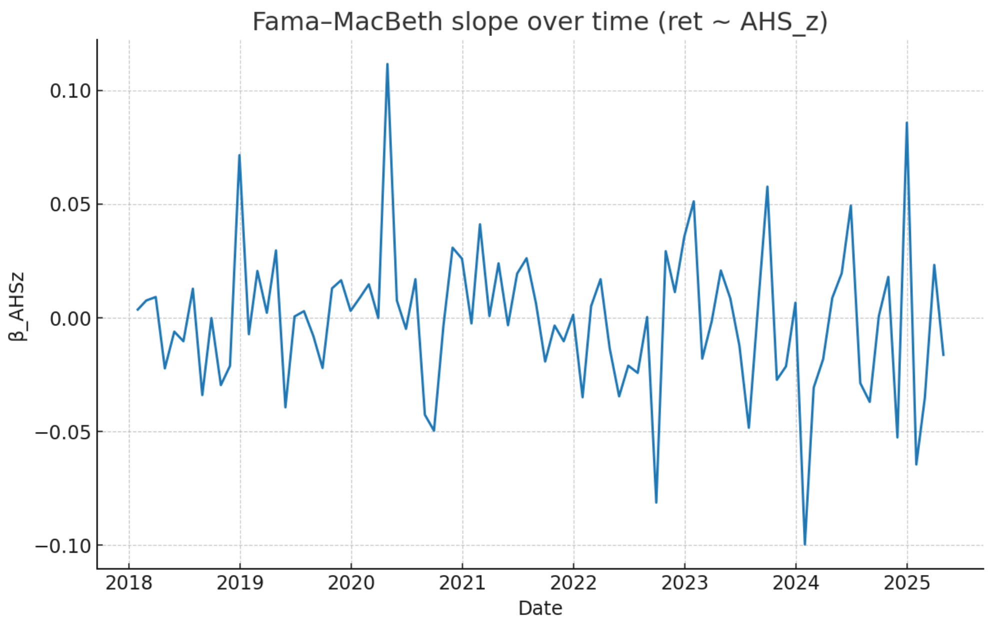

Figure 11: Fama-MacBeth Beta Coefficients Over Time

Disclaimer: Return linkage analysis is provided for descriptive context only; no trading performance is claimed.

7 Use-Cases and Practical Significance

7.1 Operational Screening and Benchmarking

The AHS+ framework provides systematic carrier comparison across multiple operational dimensions. Equity research analysts can utilize composite scores to identify airlines exhibiting superior unit economics relative to industry peers, controlling for fuel price exposure and market timing effects.

Cross-sectional rankings facilitate rapid screening of investment universes, while time-series analysis enables identification of operational inflection points that may precede fundamental performance changes.

7.2 Risk Monitoring and Attribution

The attribution framework enables decomposition of health score changes into efficiency and revenue components. This capability proves particularly valuable during periods of industry stress, allowing analysts to distinguish carriers facing temporary demand shocks from those experiencing structural operational deterioration.

Monthly updating provides timely signals for risk management applications, while the fuel-neutral specification reduces false signals from commodity price movements.

7.3 Equity Research Integration

Coverage-aware scoring accommodates varying data availability across carriers, enabling inclusion of smaller operators and private companies in comparative analyses. The standardized metrics facilitate integration with traditional valuation frameworks while providing operational context often missing from purely financial approaches.

7.4 Regulatory and Policy Analysis

The comprehensive DOT data integration supports regulatory analysis and policy impact assessment. Changes in health scores following regulatory modifications, route authority changes, or merger activity provide quantitative measures of operational impact.

8 Limitations

8.1 Data Coverage and Reporting Dependencies

The framework relies on DOT reporting requirements, limiting coverage to U.S. carriers and creating potential bias toward larger operators with more comprehensive reporting. International carriers and certain regional operators may have limited data availability.

Network pillar metrics appear in only one month across the sample period, severely constraining statistical inference for this component. Future enhancements require expanded network data collection or alternative proxy variables.

8.2 Return Linkage Limitations

Equity return associations show mixed and generally insignificant results, confirming that operational health metrics should not be interpreted as trading signals. The framework's primary value lies in operational assessment rather than market timing or security selection.

Market efficiency considerations suggest that publicly available operational metrics may be rapidly incorporated into equity prices, limiting their utility for generating excess returns.

8.3 Model Specification Constraints

The linear composition approach may not capture non-linear interactions between operational components. More sophisticated machine learning approaches could potentially improve composite construction, though at the cost of interpretability.

Stage-cost modeling relies on simplified log-linear relationships that may not adequately capture network complexity effects, aircraft mix impacts, or seasonal operational variations.

9 Conclusion

The Airline Health Score (AHS+) provides a comprehensive framework for evaluating U.S. airline operational and financial health through systematic integration of multiple data sources and standardized cross-sectional comparison. Statistical validation demonstrates strong relationships between constituent components and composite scores, with coverage-aware variants successfully addressing missing data bias while preserving ranking consistency.

The fuel-neutral approach effectively isolates operational efficiency from commodity price exposure, while the attribution framework enables decomposition of health changes into interpretable components. Extended pillars provide incremental information content, though network data limitations constrain this component's current utility.

Return linkage analysis confirms the framework's primary application in operational assessment rather than trading signal generation. The methodology offers practical value for equity research screening, risk monitoring, and benchmarking applications across the airline sector.

Future enhancements should focus on expanding network data availability, investigating non-linear composition approaches, and extending coverage to international carriers through alternative data sources.

10 Appendix

Required data for complete stage-cost documentation:

- Monthly coefficient estimates (α_t, β_t) for log-log stage-cost models
- Model fit statistics (R-squared, residual standard errors) by month
- Route-level regression diagnostics and outlier analysis
- Carrier residual backfill methodology validation

10.1 File Manifest

Primary data sources accessed:

- DOT DB1B: Origin-Destination Survey (yields, pricing)
- DOT T-100: Domestic Segment (capacity, traffic, stage lengths)
- DOT Form 41: Airline Financial Reports (detailed cost components)
- Return data: Yahoo Finance integration for publicly traded carriers

Output files generated:

- Component scores: AHS_core_z, growth_z, network_z, bs_z
- Composite scores: AHS_plus, AHS_plus_covaware, AHS_fuel_neutral
- Statistical summaries: Monthly cross-sectional correlations, return linkage metrics
- Attribution components: Efficiency and revenue decompositions