JOURNAL OF MARITIME RESEARCH

Vol XXIII. No. II (2026) pp 475–483

ISSN: 1697-4040, www.jmr.unican.es

Forecasting Oman's Blue Economy Trade: A Sector-Disaggregated Comparison of Five Small-Sample Forecasting Methods with Structural-Break Adjustment

Meena Madhavan[1,*], Mohammed Ali Sharafuddin[2,**], Abebe Ejigu Alemu[1], Faiza Kiran[1], Prabhu Mannadhan[1]

ARTICLE INFO

Article history:
Received 21 Jan 2026;
in revised form 15 Feb 2026;
accepted 28 Mar 2026.

Keywords:
blue economy, trade forecasting, ARIMA, Bayesian structural time series, exponential smoothing, forecast combination, Oman, small-sample forecasting.

JEL codes:
C53, F17, Q22

ABSTRACT

This study forecasts export and import values across four sectors of Oman's blue economy, fisheries and seafood processing, coastal and maritime tourism, maritime freight transport, and port services and logistics, using UNCTADstat trade data from 2005 to 2024. Two forecasting approaches, autoregressive integrated moving average (ARIMA) and Bayesian structural time series (BSTS), are compared against each other, against a simple equal-weight combination of both, and against two standard small-sample benchmarks, a random walk with drift and exponential smoothing (ETS), with accuracy measured throughout by genuine out-of-sample rolling-origin backtesting. Structural-break dummies for the COVID-19 disruption and the 2023 to 2024 Red Sea shipping crisis are entered as regressors where the affected sector and years overlap, and the fisheries sector is additionally cross-checked against raw UN Comtrade goods trade (HS codes 03 and 16), which runs on average 15 per cent below UNCTADstat's own ocean-adjusted figure, a gap attributable to UNCTAD's own coefficient adjustment for ocean-based activity. That same Comtrade data extends back to 2000, twelve years before UNCTADstat's own fisheries table, and a robustness check on this longer series confirms the fisheries sector's preferred method with five times as many backtest origins as the headline result alone provides. Method preference varies by sector: ARIMA wins one series, BSTS wins two, ETS wins three, and the random walk with drift wins one, with the equal-weight combination winning none outright and no method dominating across Oman's blue economy. A Holm-corrected Diebold-Mariano test was applied to all nine method rankings, the seven headline series plus the two extended fisheries checks, comparing each winner against its runner-up. None of the nine rankings reached statistical significance. A short annual backtest therefore offers limited power to distinguish between competing forecasting methods.

1. Introduction

Time series forecasting method comparisons, ARIMA against BSTS in particular, are well established in applied economics and logistics research, with both methods documented individually in the methodological literature (Hyndman and Khandakar, 2008; Scott and Varian, 2014). ARIMA-based forecasting is also established specifically within Gulf and Middle Eastern economic applications (Youssef et al., 2021), though, to this study's knowledge, no such application yet disaggregates by sector for a single small open economy the way this paper attempts. An earlier study (Madhavan et al., 2023) compared ARIMA and BSTS forecasts for the Indian airline industry, fitting both methods to ten years of monthly domestic and international passenger and cargo data across four aviation segments. That comparison found ARIMA more accurate for passenger

[1]International Maritime College Oman, National University of Science and Technology, Sohar, Oman.
[2]Qasim Ibrahim School of Business, Villa College, Malé, Maldives.
*Corresponding author: Meena Madhavan.
e-mail address: meenamadhavan@imco.edu.om.
ORCID: https://orcid.org/0000-0003-1244-5392.
**Corresponding author: Mohammed Ali Sharafuddin.
e-mail address: mohammed.ali@villacollege.edu.mv.
ORCID: https://orcid.org/0000-0001-5247-2964.

traffic and BSTS more accurate for cargo, with the gap between the two methods small in both cases. This study applies the same two methods to Oman's blue economy trade, adding three features described below.

First, forecasting is carried out separately for four distinct blue economy sectors, so that method preference can be compared across sectors. Second, structural-break dummies for dateable shocks, the COVID-19 disruption and the 2023 to 2024 Red Sea shipping crisis, are entered as exogenous regressors in both ARIMA and BSTS, giving each model an explicit and interpretable decomposition of trend and shock (Scott and Varian, 2014). Third, the underlying series are short annual counts, thirteen to twenty observations depending on sector, a different forecasting regime from the higher-frequency series used in the airline industry study. Small-sample time series forecasting behaves differently from large-sample forecasting (Hyndman and Athanasopoulos, 2021; Sharafuddin et al., 2026), and this analysis treats that plainly as a feature of the problem.

The blue economy, broadly defined as economic activity linked to oceans, seas, and coasts while preserving marine ecosystem health (World Bank and United Nations Department of Economic and Social Affairs, 2017), adds a substantive motivation of its own in the Oman context. Gulf and Arabian Peninsula economies have their own growing blue economy literature, from bibliometric mapping of the region's research output (Alhowaish, 2025) to evidence that blue economy factors contribute to sustainable growth in GCC countries specifically (Sarwar, 2022), though neither disaggregates by sector or attempts trade forecasting. Oman's economic diversification strategy places explicit weight on fisheries, tourism, and maritime logistics as sectors outside oil and gas. Forecasting their trade performance separately, with honest uncertainty bounds, carries direct planning relevance regardless of the methodological framing above.

Section 2 describes the data sources, the sector definitions, and the forecasting and evaluation methods. Section 3 presents the descriptive, accuracy, and forecast results. Section 4 discusses the findings and their limitations. Section 5 concludes.

2. Data and methods

2.1. Data sources

Trade values were obtained from two UNCTADstat analytical datasets for Oman (UNCTAD, 2026a,b), downloaded directly per Flow and Product/Category combination, covering export and import flows separately for each product or service category. The first, *Ocean goods: Bilateral trade by product group* (values in current US dollars, thousands), covers Marine fisheries, aquaculture and hatcheries, and Seafood processing, 2012 to 2024 (2025 unpublished at the time of extraction). The second, *Ocean services: Trade* (values in current US dollars, millions), covers Marine and coastal tourism, Maritime transport and related services (freight), and Port services, related infrastructure services and logistical services, 2005 to 2024 (2025 unpublished). Because the two tables report in different units, thousands of US dollars for Ocean goods and millions for Ocean services, both were converted to a common plain US dollar basis before any sector definition or model was built on them.

Marine fisheries and Seafood processing were summed into a single Fisheries and seafood processing sector. Both components are present for Oman in every year 2012 to 2024, so the sum carries no hidden composition change. Maritime freight transport and Port services and logistics were deliberately kept as two separate series. Freight has no published Oman figure for 2006 to 2017, only 2005 and 2018 to 2024 are available, while port services is complete across the full period. Summing them would have meant the combined series reflected port services alone for 2006 to 2017 and freight-plus-port from 2018 onward, a composition change that would read as trade growth in the series without being one. They are therefore reported and forecast as two separate sectors.

2.2. Series continuity

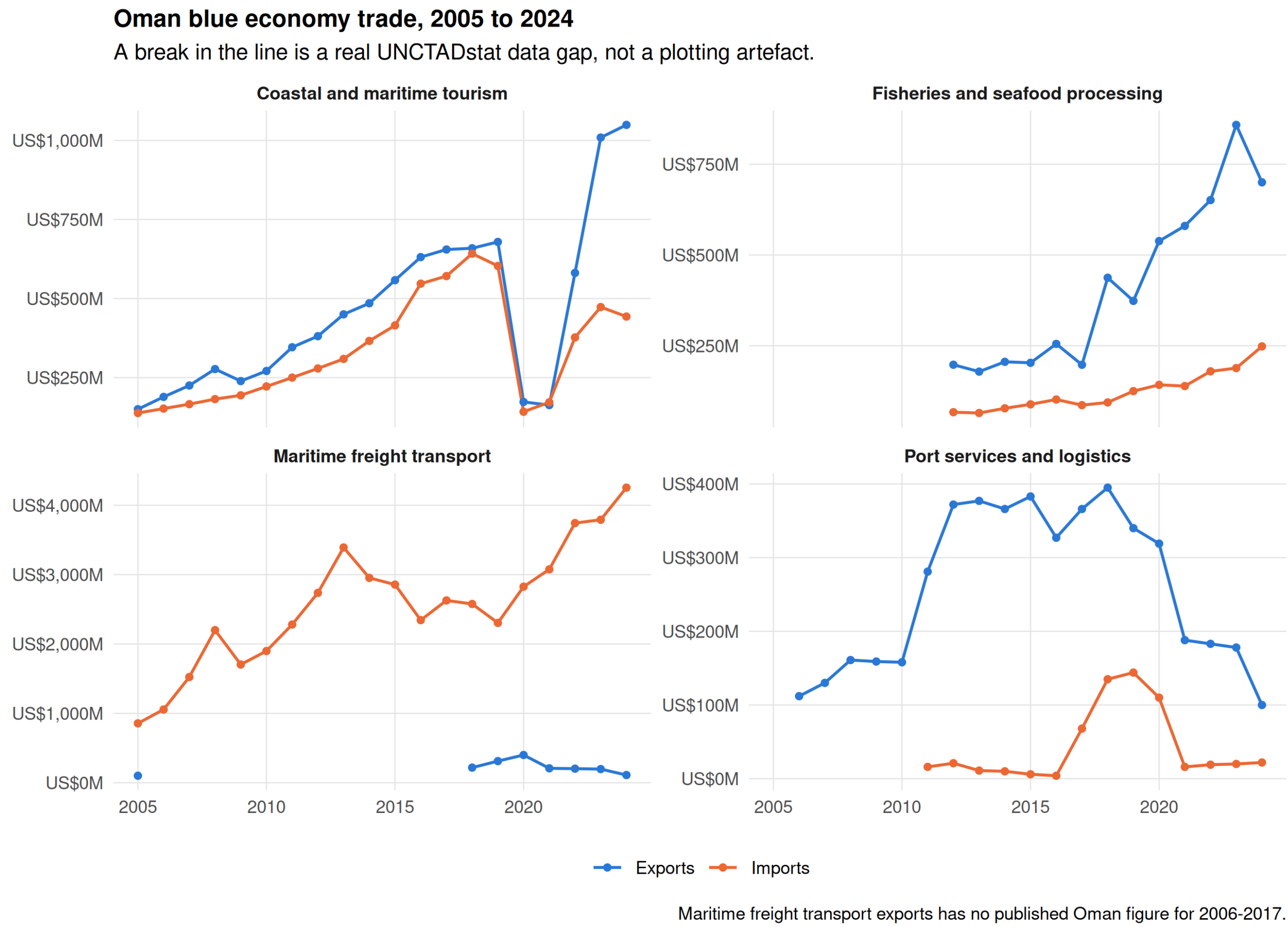
The UNCTADstat series are sometimes discontinuous. Where a series had missing years, it was restricted to its longest run of genuinely consecutive calendar years before any model was fit. Sorting a table by year and treating adjacent rows as adjacent years, without checking the actual gap between them, was rejected as a method. Maritime freight transport exports is the one series materially affected: the isolated 2005 observation was dropped, leaving seven consecutive years (2018 to 2024) for modelling, which is short enough that its rolling-origin backtest was infeasible (Section 2.4). Therefore, the full-sample forecast should be read with that limitation in mind. Figure 1 shows the full historical series for all eight sector-flow combinations, with this gap left visible as a real break in the line.

2.3. Forecasting methods

Two structural forecasting approaches were fitted to each of the eight sector-flow series independently. The first, ARIMA, was fitted via the forecast package's auto.arima() (Hyndman and Khandakar, 2008), with COVID-19 (2020 to 2021) and Red Sea shipping crisis (2023 to 2024, Maritime freight transport and Port services and logistics only) dummies entered as exogenous regressors where the series has variation on them. The second, BSTS, is a local linear trend state-space model (Scott and Varian, 2014), with the same shock dummies entered as a static regression component alongside the trend. A third approach, an equal-weight combination of the ARIMA and BSTS point forecasts at each origin, was also included, since forecast combination routinely outperforms either individual method in the applied forecasting literature (Bates and Granger, 1969). Only the simple equal-weight scheme is used here. Variance-minimizing or regression-based combination weights (Timmermann, 2006; Wang et al., 2023) were left untested, a scope limitation returned to in Section 4.1.

Two further methods were included as benchmarks, standard practice in the small-sample forecasting literature and both computable without fitting a structural model. A random walk with drift extrapolates each series' own average historical change

Figure 1: Oman blue economy trade, 2005 to 2024, real UNCTADstat series. A break in the line marks a real data gap (Maritime freight transport exports, 2006 to 2017), the result of missing UNCTADstat coverage.



forward one step at a time. Exponential smoothing (ETS) fits an error-trend-seasonal state-space model chosen automatically by its own information-criterion selection, via the `forecast` package's `ets()` function (Hyndman and Khandakar, 2008). Neither benchmark uses the structural-break dummies, since both are univariate extrapolation methods by design. Including them tests whether ARIMA, BSTS, and their combination add forecasting value over a method that assumes nothing about the shocks or structure driving each series.

2.4. *Out-of-sample evaluation*

Accuracy was evaluated using a rolling-origin, expanding-window, one-step-ahead backtest. For each origin year with at least ten years of prior data, a model was fit on data up to that origin and used to forecast the following year only, which was then compared against the actual value, giving genuine out-of-sample accuracy. All five methods were evaluated on identical origins and actual values within each series, so accuracy is directly comparable across methods. Series with fewer than eleven years of usable data after the continuity restriction (Maritime freight transport exports, seven years) lack enough data to support this backtest and are reported with a full-sample forecast only.

2.5. *Statistical significance of method rankings*

Table 3 and Table 5 rank methods by point-estimate RMSE, but with as few as three backtest origins for some series, a lower RMSE alone is insufficient to establish that one method is more accurate than another. A Diebold-Mariano test (Diebold and Mariano, 1995) was run for every backtestable series, winner against runner-up, on the squared-error loss differential from the same backtest origins reported in Table 3. The `forecast` package's `dm.test()` implementation applies the Harvey, Leybourne and Newbold (1997) small-sample correction by default, scaling the test statistic and referring it to a t-distribution with $n - 1$ degrees of freedom instead of the asymptotic normal (Harvey et al., 1997), the appropriate variant given how few origins some of these series provide. Nine tests were run in total, seven for the headline comparison and two more for the extended fisheries series (Section 2.7), so a Holm correction is also applied across all nine to control the family-wise error rate, since testing this many comparisons at an uncorrected 10 per cent threshold would be expected to produce close to one false positive by chance alone even if no real ranking difference existed anywhere in the data. Both the raw and Holm-adjusted results are given in full in Section 3.3 alongside the rankings they qualify.

2.6. Independent cross-check

Since Marine fisheries and Seafood processing map onto real Harmonized System (HS) goods codes, this sector was independently cross-checked against raw UN Comtrade bilateral trade data (United Nations Statistics Division, 2026) (HS 03, fish and crustaceans, molluscs and other aquatic invertebrates, and HS 16, prepared or preserved fish), fetched directly via the Comtrade API. Comtrade's raw HS-code total runs below UNCTADstat's ocean-adjusted figure in most years (mean difference −15.1 per cent, median −19.1 per cent, range −45.3 to +29.8 per cent across 26 overlapping year-flow observations), consistent with UNCTADstat's coefficient adjustment for ocean-based activity within broader HS-code trade. The two sources agree in order of magnitude and broad trend across every overlapping year (Figure 2). This is treated as a check on how reasonable the fisheries series looks. The two sources are built on different methodologies, so agreement in magnitude and trend is the relevant test, and neither source is treated here as the correct one against which the other is judged.

2.7. Extended-history robustness check

UN Comtrade's raw HS 03 and HS 16 data for Oman is available from 2000, twelve years earlier than UNCTADstat's Ocean goods table, which begins in 2012 for this sector (Section 2.1). This complete 2000 to 2024 series, kept as one consistent methodology throughout and never spliced with UNCTADstat's ocean-adjusted figures at 2012, since a splice would create a level discontinuity that reads as trade change without being one, was used to rerun the rolling-origin backtest for Fisheries and seafood processing exports and imports. This extends the number of backtest origins from three to fifteen per flow. UNCTADstat's ocean-adjusted figure remains this study's headline result for this sector, and the extended series serves only as a robustness check, since the other three sectors remain on UNCTADstat throughout, and this sector alone using a different data source would itself be an inconsistency. One further note: 2003 exports (US$0.25 million) is anomalously low against every neighbouring year (US$81 million in 2002, US$106 million in 2004), most likely a Comtrade reporting gap for Oman that year given the sharp recovery either side of it. It is kept in the series unaltered, consistent with this study's practice of leaving real data gaps uninterpolated, and affects one training-window observation and one origin's forecast at most.

3. Results

3.1. Descriptive overview

Table 1 summarises the eight sector-flow series. Fisheries and seafood processing is the shortest series (13 years, 2012 to 2024). Coastal and maritime tourism and Port services and logistics both show a sharp COVID-era contraction followed by recovery, visible in Figure 1.

3.2. Full-sample model specifications

Table 2 reports the ARIMA order and ETS model form selected for each series' full-sample fit, for transparency and reproducibility. ARIMA orders are mostly low (differencing alone or no autoregressive or moving-average terms beyond the structural-break regressors), which is consistent with series this short leaving little room for a richer model to be estimated reliably. BSTS convergence was checked via effective sample size and a Geweke diagnostic on the observation-noise variance chain, post burn-in. Six of the eight full-sample models converged acceptably ($|z| < 2$). Coastal and maritime tourism exports and Maritime freight transport exports showed $|z| > 3$, indicating the chain fell short of full stabilisation at the *niter* used here. Maritime freight transport exports already carries no backtest and an indicative-only forecast (Section 2.2), so this adds a second, independent reason for caution on that series specifically. Coastal and maritime tourism exports is backtestable, and ARIMA is the preferred method there in any case (Table 3), so the convergence concern leaves the reported preferred method for that series unaffected, though BSTS's own RMSE on that row should be read with this in mind.

3.3. Out-of-sample accuracy by sector and method

Table 3 reports rolling-origin one-step-ahead RMSE by method and series, visualised in Figure 3. Two small-sample benchmarks, a random walk with drift and ETS, are included alongside the three headline methods to test whether the structural models and their combination add forecasting value over simple extrapolation. Port services and logistics imports carries a MAPE above 200 per cent in the full output tables, a known problem with that statistic when actual values are close to zero in some origin years. The underlying forecasts remain reasonable in accuracy despite this problem. RMSE and MAE are the more informative statistics for that series, and Table 3 reports them instead.

3.4. Preferred method by sector and flow

Method preference varies across sectors, and widening the comparison to five methods sharpens this picture further. ARIMA is preferred for coastal and maritime tourism exports. ETS is preferred for coastal and maritime tourism imports and for both port services and logistics flows. BSTS is preferred for both fisheries and seafood processing flows. The random walk with drift is preferred for maritime freight transport imports, ahead of every structural model tested on that series, including the equal-weight combination. The equal-weight combination, included on the strength of a general finding in the forecast-combination literature (Bates and Granger, 1969), wins outright on none of the seven backtestable series in this comparison. No single method wins consistently, the result this study's sector-disaggregated design was built to surface.

These rankings are point estimates. Table 4 reports a Diebold-Mariano test (Diebold and Mariano, 1995; Harvey et al., 1997),

Figure 2: Fisheries and seafood processing: UNCTADstat's ocean-adjusted figure against raw UN Comtrade HS 03 + HS 16 trade, Oman, by flow.

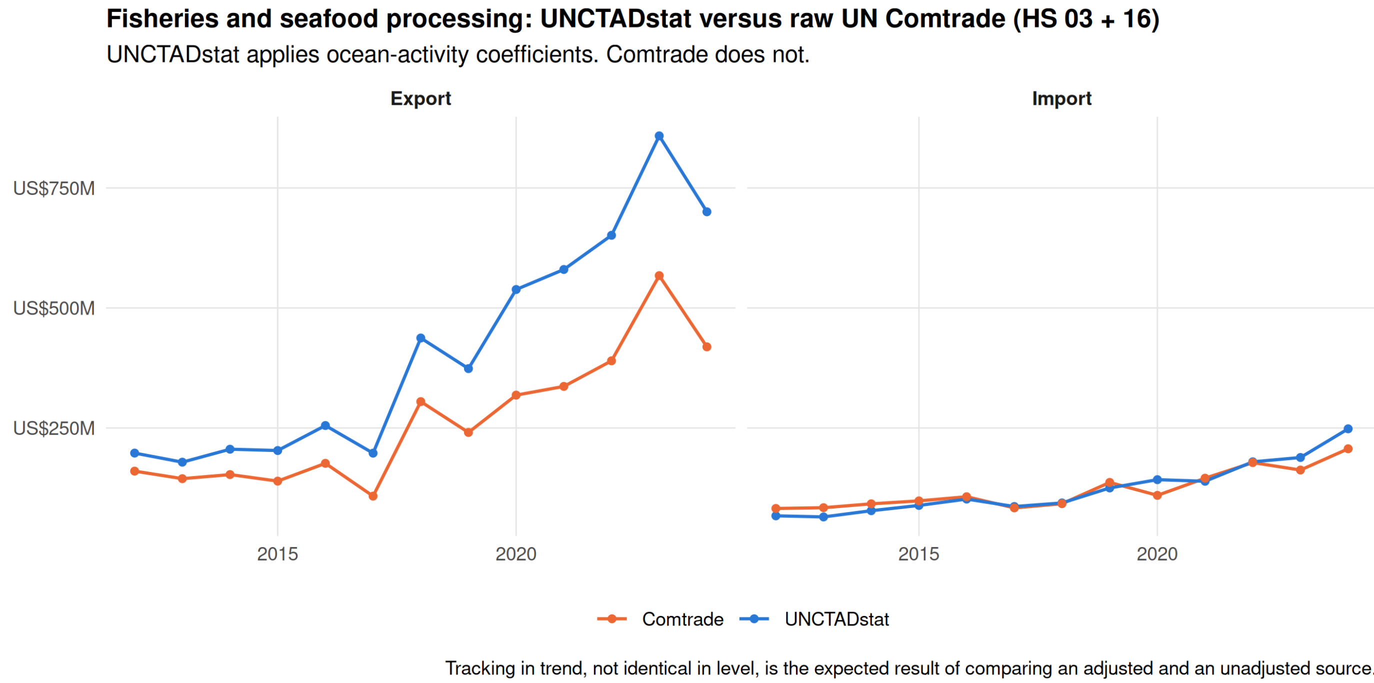


Table 1: Descriptive summary of the eight sector-flow series

Sector	Flow	Years	N	Latest (US$M)
Coastal and maritime tourism	Exports	2005–2024	20	1,081.4
Coastal and maritime tourism	Imports	2005–2024	20	444.0
Fisheries and seafood processing	Exports	2012–2024	13	837.0
Fisheries and seafood processing	Imports	2012–2024	13	246.2
Maritime freight transport	Exports	2005–2024[a]	8	211.0
Maritime freight transport	Imports	2005–2024	20	4,276.0
Port services and logistics	Exports	2006–2024	19	111.0
Port services and logistics	Imports	2011–2024	14	261.0

[a] Exports has a real gap, 2006–2017. Only 8 of the 20 nominal years have published data. Source: UNCTADstat. See Section 2.2.

winner against runner-up, for every backtestable series, alongside a Holm-adjusted p-value that accounts for all nine tests run in this analysis (the seven here plus the two more in Section 3.4). Maritime freight transport imports (random walk with drift against BSTS) is the only series significant at the uncorrected 10 per cent level ($p = 0.073$), but it fails to survive the Holm correction ($p_{\text{Holm}} = 0.657$). None of the nine rankings tested in this study, across both the headline comparison and the extended fisheries check, is statistically distinguishable from its runner-up once multiple testing is accounted for. The rankings themselves are genuine differences in point-estimate RMSE, but on three to fifteen backtest origins this study cannot say any winner is reliably more accurate than its closest competitor. This is a direct consequence of the annual, small-sample regime the Introduction already states as a defining feature of the problem, and this paper reports that uncertainty plainly, without collapsing it into false precision.

3.5. *Extended-history robustness check for fisheries*

Table 5 reports the same five-method comparison for Fisheries and seafood processing on the longer, Comtrade-based 2000 to 2024 series described in Section 2.7. BSTS remains the preferred method for both exports and imports, now backed by fifteen backtest origins per flow, up from three, a substantially larger base for that conclusion than the headline UNCTADstat-based result alone provides. The Diebold-Mariano test against the runner-up (Combination in both flows) gives $p = 0.126$ for exports and $p = 0.197$ for imports, tighter than the headline three-origin result, though neither survives the Holm correction reported in Table 4 either ($p_{\text{Holm}} = 1.000$ for both). This is consistent with Section 3.3's wider point that a handful of backtest origins rarely settles a method ranking outright, even when that handful grows from three to fifteen. Excluding the anomalous 2003 export observation (Section 2.7) as a check leaves BSTS the winner on the remaining 2004–2024 window (RMSE 96.3 against ARIMA's 102.5), confirming the ranking

Table 2: Full-sample model specifications by series

Sector	Flow	ARIMA order	ETS form
Coastal and maritime tourism	Exports	ARIMA(0,1,0)	ETS(M,A,N)
Coastal and maritime tourism	Imports	ARIMA(0,1,0)	ETS(M,A,N)
Fisheries and seafood processing	Exports	ARIMA(0,1,0)	ETS(A,N,N)
Fisheries and seafood processing	Imports	ARIMA(0,1,0)	ETS(M,N,N)
Maritime freight transport	Exports	ARIMA(0,0,0)	ETS(M,N,N)
Maritime freight transport	Imports	ARIMA(0,1,0)	ETS(M,A,N)
Port services and logistics	Exports	ARIMA(0,0,0)	ETS(A,N,N)
Port services and logistics	Imports	ARIMA(0,0,0)	ETS(A,N,N)

ARIMA orders include the structural-break dummies as regressors where applicable (Section 2.3). The (p, d, q) shown is the ARIMA part fitted to the regression residuals. ETS form follows the (error, trend, season) notation of Hyndman and Athanasopoulos (2021). Source: Authors.

Table 3: Rolling-origin out-of-sample RMSE (US$M) by method

Sector	Flow	ARIMA	BSTS	Combination	Naive (drift)	ETS
Coastal and maritime tourism	Exports	**221.2**	233.5	225.9	252.3	274.6
Coastal and maritime tourism	Imports	195.7	199.6	197.3	177.3	**176.2**
Fisheries and seafood processing	Exports	273.3	**157.6**	192.1	157.7	158.5
Fisheries and seafood processing	Imports	43.7	**31.2**	37.1	33.8	49.5
Maritime freight transport	Imports	476.5	436.6	442.8	**376.2**	560.6
Port services and logistics	Exports	66.3	67.2	64.4	69.6	**59.9**
Port services and logistics	Imports	51.2	58.2	47.7	52.3	**47.0**

Lowest RMSE per row is shown in bold. See Section 3.3 for the sector-by-sector preferred method. Maritime freight transport exports omitted, insufficient years for a backtest (Section 2.4). Source: Authors.

stays independent of that single value, though the ranking itself still falls short of statistical significance.

3.6. *Five-year forecasts*

Figure 4 shows the five-year-ahead point forecasts for all five methods, across all eight series. ARIMA, BSTS, the random walk with drift, and ETS each carry a genuine 95 per cent interval of their own, unlike the equal-weight combination. Plotting four overlapping interval bands across eight small panels made the figure harder to read without adding useful information, so Figure 4 shows point-forecast lines only, and the full interval values for every method are reported in the accompanying output files. The combination forecast has no interval of its own, since an equal-weight average of two point forecasts carries no probability model of its own.

4. Discussion

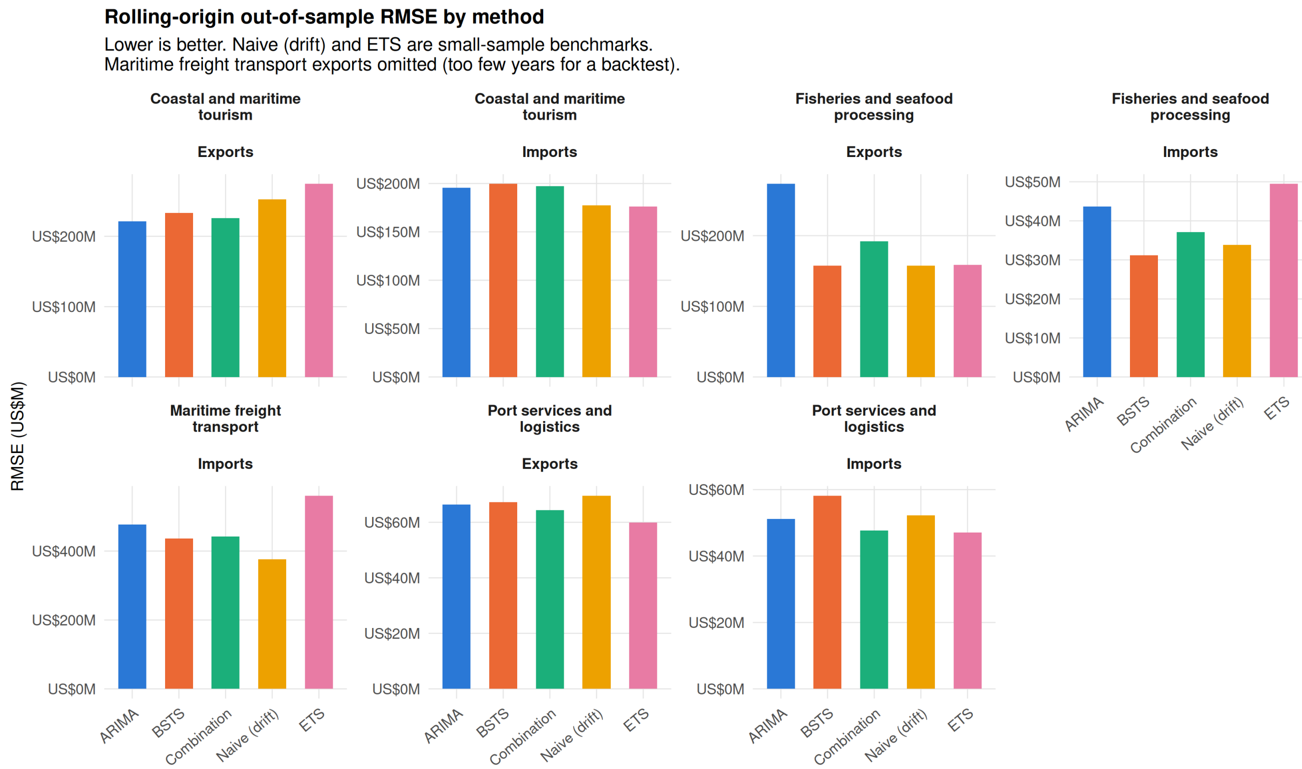
The absence of a single dominant method across sectors is consistent with the small-sample forecasting literature's general caution against treating any one time-series method as universally superior (Hyndman and Athanasopoulos, 2021; Sharafuddin et al., 2026). With thirteen to twenty annual observations per series, the data simply lack enough information to separate ARIMA's, BSTS's, ETS's, and the random walk with drift's underlying assumptions sharply, and which method wins can depend on the particular shock pattern and trend shape a given sector happened to experience. Maritime freight transport exports, restricted to seven usable years by a real UNCTADstat data gap for 2006 to 2017, is the clearest illustration of that data scarcity. Its full-sample forecast should be read as indicative only, since no backtest-based accuracy claim can honestly be made for it.

The random walk with drift's win on Maritime freight transport imports is the closest this study comes to direct evidence for that caution, and it is the one series where Section 3.3's Diebold-Mariano test is significant before correcting for the nine tests run in total ($p = 0.073$), though it fails to survive that correction ($p_{\text{Holm}} = 0.657$). This is the largest-value series in the dataset, and across all ten backtest origins a benchmark that assumes nothing beyond the series' own average historical change forecasts more accurately than ARIMA, BSTS, their combination, or ETS, even if that gap cannot be called statistically reliable once multiple testing is accounted for. The structural models still win on five of the other six backtestable series, so this single result offers no support for dropping them.

The result reflects more than model complexity alone, however. Refitting ARIMA on this series without the COVID - 19 shock dummy lowers its RMSE from US$476.5 million to US$362.5 million, below the random walk with drift's US$376.2 million. BSTS is far less sensitive to the same change (US$436.6 million with the dummy, US$431.4 million without). The shock dummy, estimated on limited pre- and post-shock variation within a training window that grows one year at a time, degraded ARIMA's accuracy on this series more than it helped, so the

Figure 3: Rolling-origin out-of-sample RMSE by method (ARIMA, BSTS, Combination, Naive with drift, ETS), all seven backtestable sector-flow series. Bars are also labelled by method on the axis, so identity is readable independent of colour.

Source: Authors.



naive benchmark's win over the ARIMA model reported in Table 2 partly reflects that specific estimation cost, more than any general advantage of simplicity. The structural - break dummies still stay in the headline models, since a single series' backtest offers no grounds for redesigning the method applied uniformly across all eight, but the finding qualifies the complexity claim: what beat ARIMA here was, in part, a specific regressor's small-sample estimation cost.

The structural-break dummies were included as regressors. Their coefficients are reported without a claim of statistical significance. Neither ARIMA's nor BSTS's regression coefficients are interpreted here as a causal claim about the effect size of COVID-19 or the Red Sea crisis. The fitted models' primary purpose is forecasting alone, and a formal significance test of shock coefficients fell outside the scope of this analysis.

The Comtrade cross-check supports the fisheries and seafood processing series as a reasonable reading of Oman's actual trade, since an independent source built on different methodology shows the same order of magnitude and trend. It also makes clear that choosing between the two sources is a matter of scope: raw HS-code trade against a deliberately ocean-adjusted composite, with neither source more accurate than the other in principle. Future work extending this comparison to the remaining three sectors would require UNCTAD's own ocean-classification coefficients, which lie beyond independent reconstruction from Comtrade alone.

For policymakers, three points from these results carry direct planning relevance. First, the finding that no single method dominates argues against any one government forecasting unit standardising on a single technique across every blue economy sector, since Table 2 and the Diebold-Mariano results together show this would trade real accuracy for administrative convenience, with the specific method that performs best differing by sector. Second, the five-year forecasts in Figure 4 point to continued growth in coastal and maritime tourism, fisheries and seafood processing exports, and port services and logistics, consistent with the ambitions of Oman's diversification strategy, but the width of the intervals reported in the underlying output files (Section 5) is a real part of that forecast that planning should carry forward. Planning that assumes only the central point forecast risks being wrong-footed by ordinary annual variation on series this short. Third, Maritime freight transport imports, the largest-value series in the dataset, is also the one series where this study's own accuracy comparison is least settled: even its winning method's advantage fails to survive the correction reported in Table 4, so planning decisions resting specifically on this series' forecast should treat it as the most provisional of the eight.

4.1. Limitations

The fisheries and seafood processing series begins in 2012, eight years after the rest of the dataset, because UNCTADstat's Ocean goods table itself extends no earlier for Oman, a real boundary of that specific table: Section 2.7's extended-history

Table 4: Diebold-Mariano test, winner against runner-up, all seven backtestable series, with Holm correction across all nine tests (including the two in Table 5)

Sector	Flow	Winner	Runner-up	Origins	DM statistic	p-value	Holm p
Coastal and maritime tourism	Exports	ARIMA	Combination	10	−0.95	0.183	1.000
Coastal and maritime tourism	Imports	ETS	Naive (drift)	10	−0.12	0.455	1.000
Fisheries and seafood processing	Exports	BSTS	Naive (drift)	3	−0.03	0.491	1.000
Fisheries and seafood processing	Imports	BSTS	Naive (drift)	3	−1.40	0.148	1.000
Maritime freight transport	Imports	Naive (drift)	BSTS	10	−1.59	0.073	0.657
Port services and logistics	Exports	ETS	Combination	9	−0.94	0.188	1.000
Port services and logistics	Imports	ETS	Combination	4	−0.47	0.334	1.000

One-sided test (winner's squared error less than runner-up's), with the Harvey-Leybourne-Newbold small-sample correction. No row is significant after the Holm adjustment. Maritime freight transport exports omitted, no backtest (Section 2.4). Source: Authors.

Table 5: Fisheries and seafood processing, extended series (Comtrade, 2000–2024): rolling-origin out-of-sample RMSE (US$M) by method, 15 backtest origins per flow

Flow	ARIMA	BSTS	Combination	Naive (drift)	ETS
Exports	94.8	**82.5**	87.1	87.6	88.8
Imports	26.2	**19.3**	21.4	22.1	30.1

Lowest RMSE per flow is shown in bold. This series uses raw Comtrade HS 03 + HS 16 trade throughout, so its RMSE values are on a different level basis from Table 2's fisheries row and should be compared within this table only. Source: Authors.

robustness check, drawing on Comtrade's longer coverage back to 2000, confirms BSTS remains the preferred method for this sector regardless. Maritime freight transport exports' seven-year usable window is likewise a real constraint set by a UNCTADstat gap, and no equivalent longer-history check was available for that sector, since Maritime freight transport maps onto no single clean HS code the way fisheries does. Neither gap was interpolated or otherwise filled.

The equal-weight combination was the only combination scheme tested. Variance-minimizing or regression-based weights (Timmermann, 2006; Wang et al., 2023) might outperform a flat average on some series, particularly given the combination's RMSE sits within 10 per cent of the winning method on three of Table 3's seven rows without matching it, but this remains untested here and is left for future work. The random walk with drift and ETS were fitted without the structural-break dummies by design, so the comparison between them and ARIMA or BSTS mixes two things, model type and shock-adjustment, without isolating either one. Section 4's finding on Maritime freight transport imports shows this confound matters in at least one case. A fuller separation, fitting ARIMA and ETS both with and without the dummies across all eight series, is left for future work here.

5. Conclusions

Forecasting Oman's blue economy trade sector by sector, using out-of-sample accuracy to compare ARIMA, BSTS, their equal-weight combination, and two small-sample benchmarks, a random walk with drift and ETS, shows that method preference varies by sector: ARIMA, BSTS, ETS, and the random walk with drift each win on a distinct subset of Oman's blue economy trade flows, and the equal-weight combination wins none outright despite its general strength in the forecast-combination literature. A Holm-corrected Diebold-Mariano test finds none of these rankings statistically distinguishable from their runner-up, a direct measure of how little a ten-origin-or-shorter backtest can resolve on its own, and a finding this paper reports plainly, without letting a single uncorrected p-value below 0.10 imply more precision than nine tests on this little data can support. This heterogeneity, together with the short, sometimes discontinuous annual series that UNCTADstat itself makes available for a small open economy like Oman, argues for reporting sector-specific forecasts and their real data limitations honestly, without collapsing them into a single blended headline number.

Acknowledgment

We thank Ministry of Higher Education, Research and Innovation (MoHERI) of the Sultanate of Oman for funding this project.

Funding

The research leading to these results has received funding from the Ministry of Higher Education, Research and Innovation (MoHERI) of the Sultanate of Oman under BFP. MoHERI Block Funding Agreement No BFP/RGP/EHR/25/384.

Author Contributions

Meena Madhavan acquired funding, contributed to conceptualisation, methodology, and writing (original draft, review and editing). Mohammed Ali Sharafuddin is responsible for data curation, formal analysis, methodology, software, and writing (original draft, review and editing). Faiza Kiran, Abebe Ejigu Alemu, and Prabhu Mannadhan contributed in review and edit the manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Generative AI Use

During the preparation of this work, the authors used Claude (Anthropic) to assist with data processing scripts and language proofing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data and Code Availability Statement

UNCTADstat data are available at `https://unctadstat.unctad.org`. UN Comtrade data are available at `https://comtradeplus.un.org` via the free API (registration required). All data and R analysis scripts are archived at `https://github.com/MohammedAliSharafuddin/oman-blue-economy-trade-forecasting` and permanently archived on Zenodo at `https://zenodo.org/records/21946958`. Analysis was run under R 4.6.0 (R Core Team, 2026), with `forecast` 9.0.2 (Hyndman and Khandakar, 2008), `bsts` 0.9.11 (Scott and Varian, 2014), `comtradr` 1.0.6 (Bochtler et al., 2026), and `dplyr` 1.2.1 (Wickham et al., 2026). The BSTS Markov chain Monte Carlo sampler used a fixed seed (2026) for every model reported. This seed does not, however, make BSTS fully deterministic end to end, and independent reruns of the archived pipeline on identical input data were found to move BSTS-derived figures, and the Combination and Diebold - Mariano statistics that depend on them, by roughly 0 to 3 per cent. Every ranking and significance conclusion reported here held across repeated runs. The specific decimal values in Tables 4 and 5 may not match a fresh run bit for bit. ARIMA, the naive-drift and ETS baselines, and the Comtrade cross-check percentages in Section 2.6 are fully deterministic

Conflicts of Interest

The authors declare no conflicts of interest.

References

Alhowaish, A. K., 2025. The blue economy in the Arabian Gulf: trends, gaps, and pathways for sustainable coastal development. Sustainability 17 (19), 8809.
DOI: `10.3390/su17198809`

Bates, J. M., Granger, C. W. J., 1969. The combination of forecasts. Operational Research Quarterly 20 (4), 451–468.
DOI: `10.1057/jors.1969.103`

Bochtler, P., Goers, H., Muir, C., 2026. comtradr: Interface with the United Nations Comtrade API. R package version 1.0.6.
URL: `https://CRAN.R-project.org/package=comtradr`
DOI: `10.32614/CRAN.package.comtradr`

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13 (3), 253–263.
DOI: `10.1080/07350015.1995.10524599`

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting 13 (2), 281–291.
DOI: `10.1016/S0169-2070(96)00719-4`

Hyndman, R. J., Athanasopoulos, G., 2021. Forecasting: Principles and Practice, 3rd Edition. OTexts, Melbourne, Australia.
URL: `https://otexts.com/fpp3/`

Hyndman, R. J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software 27 (3), 1–22.
DOI: `10.18637/jss.v027.i03`

Madhavan, M., Sharafuddin, M. A., Piboonrungroj, P., Yang, C.-C., 2023. Short-term forecasting for airline industry: the case of Indian air passenger and air cargo. Global Business Review 24 (6), 1145–1179.

R Core Team, 2026. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL: `https://www.R-project.org/`
DOI: `10.32614/R.manuals`

Sarwar, S., 2022. Impact of energy intensity, green economy and blue economy to achieve sustainable economic growth in GCC countries: does Saudi Vision 2030 matter to GCC countries. Renewable Energy 191, 30–46.
DOI: `10.1016/j.renene.2022.03.122`

Scott, S. L., Varian, H. R., 2014. Predicting the present with Bayesian structural time series. International Journal of Mathematical Modelling and Numerical Optimisation 5 (1-2), 4–23.
DOI: `10.1504/IJMMNO.2014.059942`

Sharafuddin, M. A., Jaleel, A. A., Madhavan, M., 2026. Quantitative Analysis with Small Samples: A Practical Guide for Students and Early-Career Researchers. Zenodo.
DOI: `10.5281/zenodo.20221929`

Timmermann, A., 2006. Forecast combinations. In: Elliott, G., Granger, C. W. J., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 1. Elsevier, Ch. 4, pp. 135–196.
DOI: `10.1016/S1574-0706(05)01004-9`

UNCTAD, 2026a. Ocean goods: Bilateral trade by product group, annual (analytical). Tech. rep., UN Trade and Development, UNCTADstat Data Centre.

UNCTAD, 2026b. Ocean services: Trade, annual (analytical). Tech. rep., UN Trade and Development, UNCTADstat Data Centre.

United Nations Statistics Division, 2026. UN Comtrade Database. Tech. rep., United Nations, New York.

Wang, X., Hyndman, R. J., Li, F., Kang, Y., 2023. Forecast combinations: an over 50-year review. International Journal of Forecasting 39 (4), 1518–1547.
DOI: `10.1016/j.ijforecast.2022.11.005`

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., 2026. dplyr: A Grammar of Data Manipulation. R package version 1.2.1.
URL: `https://CRAN.R-project.org/package=dplyr`
DOI: `10.32614/CRAN.package.dplyr`

World Bank and United Nations Department of Economic and Social Affairs, 2017. The potential of the blue economy: Increasing long-term benefits of the sustainable use of marine resources for small island developing states and coastal least developed countries. Tech. rep., World Bank, Washington, DC.
DOI: `10.1596/26843`

Youssef, J., Ishker, N., Fakhreddine, N., 2021. GDP forecast of the biggest GCC economies using ARIMA. MPRA Paper 108912, University Library of Munich.

Figure 4: Five-year forecasts by method, all eight sector-flow series and all five methods (ARIMA, BSTS, Combination, Naive with drift, ETS). Grey is historical.
Source: Authors.