Working paper · Internal review | v1.0 | 2026-11-20

A monthly forecast benchmark for stay-over arrivals in Saint Lucia

Seasonal persistence and a fitted trend across six annual test periods

Adrian Dunkley | info@caribbeanaiassociation.com

Abstract

This study tests whether adding a fitted trend improves monthly stay-over arrival forecasts in Saint Lucia. At the end of each year, we estimated twelve monthly counts using the preceding year's monthly arrivals and, separately, a regression with calendar-month effects and a shared linear trend. We evaluated 60 monthly forecasts per method from 2015 through 2019 and examined 2020 separately. The mean absolute monthly error before 2020 was 1,867 arrivals for the prior-year rule and 1,885 for the regression. In 2020 the errors rose to 24,666 and 24,443. The monthly source table contains two annual reconciliation differences. We retained its monthly observations and report both discrepancies. The comparison gives local teams a reproducible entry test for more complex forecasting models.

Research question

Does a fitted trend improve on last year's monthly count, and how do both methods perform when the historical pattern breaks?

Data and methods

We extracted the twelve monthly counts for each year from 2010 through 2020 from the Central Statistical Office's tourist arrivals table, updated on 16 May 2024 [1]. The Office's annual series identifies the corresponding target as stay-over arrivals [2]. We retained each published month, including the three recorded zeros in April, May and June 2020. The monthly source marks all 2021 observations as missing. We ended the sequence in 2020 rather than bridging that gap.

At the end of each year from 2014 through 2019, we issued twelve forecasts for the next year. The first method copied the count for the same month of the preceding year. The second fitted an ordinary least-squares regression to all monthly counts from 2010 through the forecast origin. It assigned each calendar month its own intercept and estimated one annual trend shared by all months. Neither method received observations from the target year. The design follows time-ordered evaluation with simple forecast benchmarks [3].

We scored the absolute difference between the forecast and the published monthly count. We pooled the 60 monthly errors for 2015–2019, weighting each month equally, and reported the twelve errors for 2020 separately. Monthly percentage error would be undefined for the three months with zero arrivals. We also checked the sum of each year's monthly counts against the total printed in the same table [1].

Results

For 2015–2019, the prior-year rule missed by an average of 1,867 arrivals per month. The fitted trend missed by 1,885, a difference of 18 arrivals. The regression performed better in 2016 and 2017. It performed worse in 2015, 2018 and 2019. Both methods tended to underpredict during this period. Mean signed errors, defined as forecast minus observation, were −1,426 for the prior-year rule and −1,295 for the fitted trend.

In 2020 the monthly mean absolute error rose to 24,666 for the prior-year rule and 24,443 for the regression. For one example, the prior-year method assigned 36,828 arrivals to April 2020. The Office recorded zero [1]. The two models had no input that identified an abrupt change in travel conditions.

The monthly counts for 2017 sum to 386,327, which exceeds the published total of 386,127 by 200. The 2019 months sum to 423,741, five above the reported 423,736. The other nine years reconcile exactly. The annual series also reports 386,127 for 2017 and 423,736 for 2019 [2]. We kept the published monthly values because the sources do not identify a month to correct.

Mean absolute monthly forecast error

arrivals per month



Each bar averages absolute errors across 60 months in 2015–2019 or 12 months in 2020. Lower values indicate smaller errors. Calculations from [1].

Discussion

The average pre-2020 difference is too small, relative to typical month-to-month errors, to justify a preference from this sample alone. A team evaluating a machine-learning forecast can compare it against both benchmarks at the same twelve monthly horizons. Reporting each year separately will show whether one unusual year accounts for most of a claimed gain.

The 2020 figures expose a practical limit of a calendar-only model. A team with dated operating information could examine whether adding flight schedules or announced travel restrictions improves forecasts. Such a test must use the schedule or announcement available on the forecast date. Later revisions would give the model information that planners did not yet have.

Limitations

This is a retrospective test with one island, five ordinary evaluation years and one disruption year. No test of statistical significance or probability interval is reported. Consecutive monthly errors can be related, and the pooled score gives equal weight to all twelve horizons.

The source release postdates the forecast dates. Earlier public releases may have contained different counts. The two reconciliation differences limit annual comparisons, and annual arrivals cannot indicate which properties, routes or workers experienced a change. The study does not estimate the causal effect of any travel policy.

Reproduce the analysis

Use the twelve month rows. Keep the Total row only for checking monthly sums. For each target year from 2015 through 2020, train on January 2010 through December of the preceding year.

For each target month, copy its most recent annual observation as the first forecast. For the second, fit monthly count on twelve calendar-month indicators and calendar year, without a separate intercept. Predict all twelve target months using this fit.

Take absolute errors from the published monthly values. Average over 2015–2019 and separately over 2020. For bias, retain the sign of forecast minus observation. Round only the reported scores.

Data table

Month	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Jan	26083	26993	25605	25899	27643	31541	30661	31649	32755	34546	33675
Feb	27867	26142	28947	27853	30135	32083	31637	33402	35536	38166	39638
Mar	29580	29536	30885	33842	34538	34595	35069	35427	41741	41674	18086
Apr	25984	29122	27399	27772	30757	31200	30104	34322	32555	36828	0
May	30349	24786	24257	26679	27676	29658	29813	31436	32277	34354	0
Jun	22993	22404	21151	24071	25268	26347	25855	31312	31318	34040	0
Jul	34186	32835	29416	28428	32100	32481	33381	38291	37844	42778	1955
Aug	29589	28429	27866	27536	28646	27822	29991	33400	34168	36342	3261
Sep	17393	16844	16687	18391	18247	17468	18169	20249	18745	21608	4122
Oct	20624	22431	22248	22385	22805	23745	24251	27967	25991	26923	7808
Nov	14741	22536	23709	25167	26933	25821	25254	29962	30422	32913	7766
Dec	26548	30346	28631	30603	33410	32147	33687	38910	41428	43569	14384
Total	305937	312404	306801	318626	338158	344908	347872	386127	394780	423736	130695

References

[1] Central Statistical Office of Saint Lucia. Tourist Arrivals Monthly 2010 to 2023. Updated 16 May 2024.

[2] Central Statistical Office of Saint Lucia. Selected Visitor Statistics, 2012 to 2023. Stay-over arrivals series.

[3] Hyndman, R. J., and Athanasopoulos, G. Forecasting: Principles and Practice, third edition. Chapter 5, The forecaster’s toolbox.

Data credits

This product was adapted from the information of The Central Statistical Office of Saint Lucia, which is licensed under the Open Licence Agreement of The Central Statistical Office of Saint Lucia. Contributors: Saint Lucia Tourist Board, Saint Lucia Air and Sea Ports Authority, Central Statistical Office.