Journal of Economics
AND MANAGEMENT SCIENTIES
https://jems.ink
ISSN 2655-1934 (print), 2655-6685 (online)

Forecasting the Feasibility of a Sales Budget Expansion Target in Shipping Logistics: Evidence from PT XYZ

Wilujeng Dewi Ariyanti[1], and Timotius FCW Sutrisno[2]*

[1,2] *Universitas Ciputra, Indonesia*

Journal of Economics and Management Scienties is licensed under a Creative Commons 4.0 International License.

ARTICLE HISTORY

Received: 14 October 26
Final Revision: 24 October 26
Accepted: 02 November 26
Online Publication: 31 December 26

KEYWORDS

Forecasting, Feasibility, Sales Budget, Expansion Target, Shipping Logistics

A B S T R A C T

Annual sales budgets in shipping logistics are often set by managerial ambition rather than evidence. This study tests whether a 100% incremental sales target is quantitatively feasible, using 36 months of revenue and operational data (June 2023 to May 2026) from PT XYZ, an Indonesian freight forwarding provider. Ten forecasting models in three groups (baseline, statistical and decomposition-based, including Prophet-type and empirical wavelet transform hybrids) were compared on a 12-month hold-out year. The chosen model was re-evaluated to predict revenue for FY2027, and a bootstrap simulation was used to determine the likelihood of achieving the USD 21.99 million target. The simple model performed best, with a MAPE of 9.62%, and none of the decomposition models surpassed the basic benchmarks. It was found that the length of the forecasting period had a greater impact on accuracy than the complexity of the model. The forecast for FY2027 revenue is USD 12.57 million, which is an increase of 14.3%, but this only meets 57% of the target. None of 30,000 simulated outcomes reached the target, which would require operational volumes to grow by 100% to 184%, against a record of 82.7%. An evidence-based committed budget of USD 12.74 million and a stretch budget of USD 13.10 million are proposed. The study links forecasting research with budget target-setting theory by measuring target difficulty before the budget year and offers a transparent procedure that firms with short data histories can replicate.

KATA KUNCI

Peramalan, Kelayakan, Anggaran Penjualan, Target Ekspansi, Logistik Pengiriman

CORRESPONDING AUTHOR

wariyanti@magister.ciputra.ac.id

DOI

10.37034/jems.v9i1.584

A B S T R A K

Anggaran penjualan tahunan di industri logistik pengiriman sering ditetapkan berdasarkan ambisi manajerial, bukan bukti. Penelitian ini menguji apakah target penjualan dengan pertumbuhan inkremental 100% layak secara kuantitatif, menggunakan 36 bulan data pendapatan dan operasional (Juni 2023 hingga Mei 2026) dari PT XYZ, penyedia jasa *freight forwarding* di Indonesia. Sepuluh model peramalan dalam tiga kelompok (*baseline*, statistik, dan berbasis dekomposisi, termasuk model tipe *Prophet* dan hibrida *empirical wavelet transform*) dibandingkan pada periode *hold-out* selama 12 bulan. Model terpilih diestimasi ulang untuk meramalkan FY2027, dan simulasi *bootstrap* mengubah ramalan menjadi probabilitas tercapainya target sebesar USD 21,99 juta. Model *naïve* merupakan model paling akurat (MAPE 9,62%), dan tidak ada model dekomposisi yang mengungguli pembanding sederhana; horizon peramalan lebih menentukan daripada kompleksitas model. Pendapatan FY2027 diramalkan sebesar USD 12,57 juta (+14,3%), atau hanya 57% dari target. Tidak satu pun dari 30.000 hasil simulasi mencapai target, yang menuntut volume operasional tumbuh 100% hingga 184%, sementara rekor historisnya hanya 82,7%. Penelitian ini mengusulkan anggaran komitmen berbasis bukti sebesar USD 12,74 juta dan anggaran *stretch* sebesar USD 13,10 juta. Penelitian ini menghubungkan riset peramalan dengan teori penetapan target anggaran dengan mengukur kesulitan target sebelum tahun anggaran, serta menawarkan prosedur transparan yang dapat direplikasi oleh perusahaan dengan riwayat data yang pendek.

1. Introduction

Sales budgeting in shipping logistics is, at its core, forecasting exercise. Revenue in international express and freight forwarding is a derived demand that follows trade flows, e-commerce activity and the stocking decisions of shippers more closely than the carrier's own sales effort [1], [2]. Any annual sales budget therefore contains an implicit forecast of this demand, and the budget is only as credible as that forecast.

Forecasting this demand has become more difficult in recent years. The COVID-19 pandemic triggered extreme swings in shipping volumes and freight rates, which were followed by an equally sudden normalization [3], [4]. Large integrators now expect little top-line growth; UPS, which serves here as the

reference business case for an integrated express carrier, reported revenue of USD 88.7 billion for 2025 and guided only about USD 89.7 billion for 2026 [5].

The Indonesian market adds further structural pressure. Logistics costs still amount to about 14.29% of national GDP [6]. In this setting, an ambitious sales budget is both a commercial aspiration and a wager against a volatile, cost-constrained environment.

PT XYZ, an Indonesian freight forwarding company that operates similarly to an integrated carrier, is currently facing this situation. Its FY2027 budget projects a 100% increase in revenue compared to FY2026, reaching USD 21.99 million, which is twice the revenue recorded in FY2026.However, recent performance does not support such a significant growth. Revenue increased from USD 8.24 million in FY2024 to USD 8.53 million in FY2025, a rise of 3.5%, and then to USD 11.00 million in FY2026, an increase of 28.8%.

The growth seen in FY2026 was primarily due to higher volume rather than higher prices. Shipped weight increased by about 83% and the number of active customers by about 51%, whereas revenue per kilogram dropped by roughly 29%. This dilution of yield implies that the engine behind FY2026 may not convert proportionally into revenue, even though an expansion budget assumes exactly that.

The gap between the target and the evidence is therefore the central managerial problem of this study. In practice, companies usually deal with these gaps by discussing and agreeing rather than by thoroughly examining the issue, which leaves unclear whether the target is achievable and, if it isn't, what level of revenue can be supported by data. Solving this issue needs two types of knowledge that are rarely used together.

The first body of knowledge is forecasting research. Large-scale comparisons show that simple, well-specified statistical methods are difficult to outperform, that complex and decomposition-based models must be tested against them out of sample, and that probabilistic forecasts are required to judge how likely an outcome is [7], [8], [9]. The second is budgeting research, which finds that targets serve both planning and motivational roles. It also finds that most targets are set to be reached with high probability, and that excessive or ratcheted targets can weaken performance [10], [11], [12], [13].

These two streams of research rarely intersect. Forecasting studies focus on accuracy without asking whether a management target can be reached, while budgeting studies examine target difficulty without measuring it against validated forecasts. In logistics, forecasting work concentrates on physical volumes such as freight tonnage or container throughput rather than on revenue budgets, and evidence from emerging-market carriers remains limited [2].

To the best of our knowledge, no prior study has combined baseline, statistical and decomposition-based forecasting models to test an explicit sales expansion target in an Indonesian shipping logistics firm. This study fills that gap by forecasting PT XYZ's FY2027 revenue and assessing whether the 100% incremental target is quantitatively feasible. Ten models in three groups are compared on monthly revenue from FY2024 to FY2026, with FY2026 held out for evaluation.

The three model groups cover a range of complexities. The baseline group contains the naïve, historical-growth and three-month moving-average methods, and the statistical group contains simple exponential smoothing, ARIMA and the seasonal naïve model. The decomposition group contains a Prophet-type additive model, Holt-Winters and their empirical wavelet transform hybrids.

Three research questions guide the analysis. RQ1 asks which approach forecasts PT XYZ's monthly revenue most accurately, measured by MAE, MAPE and forecast bias on the hold-out year. RQ2 asks what FY2027 revenue the best approach, or a combination of approaches, projects and with what uncertainty, while RQ3 asks how large the gap between the target and the forecast is and what it implies for a defensible sales budget.

The study offers three contributions. Theoretically, it links forecasting research with budget target-setting theory by defining target feasibility as the likelihood of achieving the target based on a validated forecast. Methodologically, it offers a clear, step-by-step process that companies with limited historical data can follow using a spreadsheet or open-source software, and practically, it provides a way to set realistic targets PT XYZ and similar carriers in emerging markets an evidence-based basis for negotiating and revising expansion targets.

Based on this literature, three propositions are tested. P1 states that, on a short monthly revenue series, simple statistical models forecast at least as accurately as decomposition-based models [7], [14]. P2 states that an equal-weight combination of the best models forecasts more accurately than the single best model [15], [16], and P3 states that the FY2027 target of USD 21.99 million lies above the upper bound of the forecast's prediction interval, so that it is infeasible without separately quantified growth initiatives [10], [11], [13].

2. Research Method

2.1. Research design and procedure

This study adopts a quantitative, single-case predictive design. PT XYZ is treated as an embedded case of an Indonesian shipping logistics firm whose 100% incremental FY2027 target can be tested against its own revenue history. A single-case design suits the research question because it concerns one firm's budget decision,

and the internal data needed is not available from public sources.

The procedure follows the framework in Figure 1 and consists of four stages. In the first stage, monthly revenue and operational drivers for FY2024 to FY2026 are extracted, cleaned and ordered chronologically. In the second stage (RQ1), ten models in three groups are estimated on the training period and evaluated on the FY2026 hold-out year. The third stage (RQ2) re-estimates the best model and an equal-weight combination of the top models on all 36 months to forecast FY2027 with prediction intervals. In the fourth stage (RQ3), the FY2027 forecast distribution is compared with the target. This comparison quantifies the gap, the probability of attainment and the operational growth that the target would require.

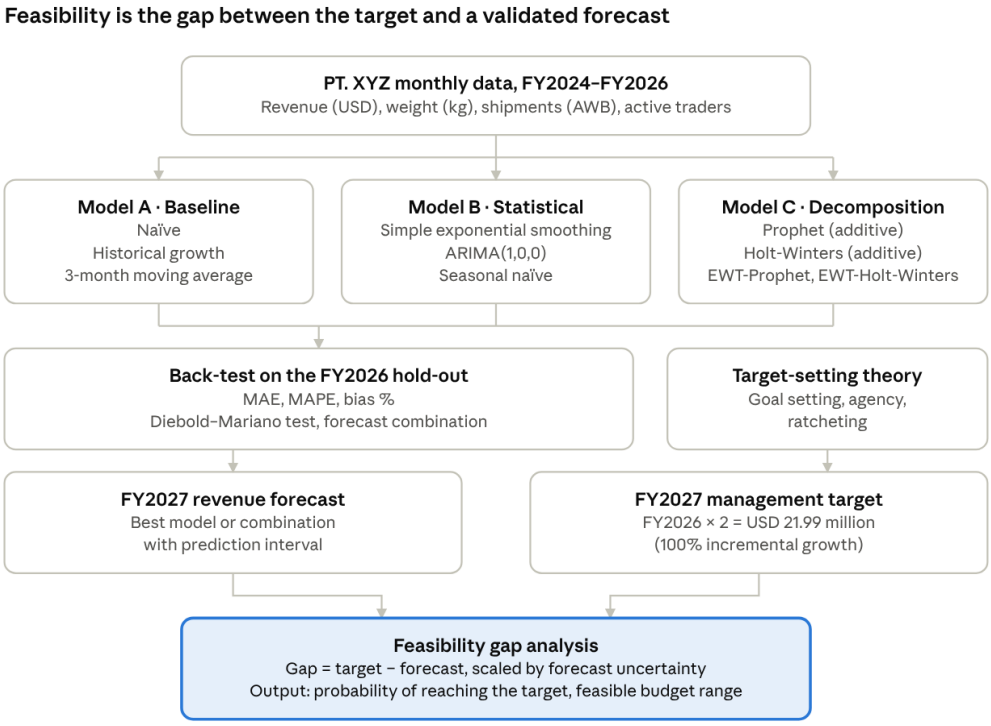


Figure 1. Conceptual framework: from data to the feasibility verdict.

Table 1 shows how the data are divided across the stages. The first two fiscal years are used for estimation and FY2026 serves as the test year. All models are then re-estimated for the full 36 months before FY2027 is forecast.

Table 1. Data partition

Period	Calendar months	Observations	Role
FY2024	Jun 2023 – May 2024	12	Training
FY2025	Jun 2024 – May 2025	12	Training
FY2026	Jun 2025 – May 2026	12	Hold-out (test)
FY2027	Jun 2026 – May 2027	12 (forecast)	Out-of-sample forecast; models re-estimated on FY2024–FY2026

2.2. Case setting, data and variables

PT XYZ is an Indonesian provider of international freight forwarding services whose business model resembles that of integrated carriers such as UPS. The company's name has been kept confidential. Its fiscal year starts in June and ends in May. For FY2027, the sales budget projects revenue to be double the actual revenue from FY2026, which is USD 21.99 million.

The data are sourced from the company's internal monthly sales performance reports. These reports track revenue and operational factors for each sales account. These records were aggregated into firm-level monthly totals. Revenue is the dependent variable, while the operational drivers are used to explain past growth and to check the feasibility of the target rather than as regressors in the forecasting models as other details can be seen on Table 2.

Table 2. Variables

Variable	Unit	Definition	Role
Revenue	USD	Monthly invoiced sales revenue	Forecast target (dependent variable)
Weight	kg	Chargeable weight shipped per month	Volume driver
Consignments	AWB count	Number of air waybills (shipments) per month	Volume driver
Active traders	Customers	Number of customers shipping in the month	Customer-base driver
Yield	USD per kg	Revenue ÷ weight	Derived price driver
Revenue per consignment	USD per AWB	Revenue ÷ consignments	Derived price driver

The months were arranged chronologically from June 2023 to May 2026 to form a single series of 36 observations, and the annual total columns of the source report were excluded. Each fiscal-year total was reconciled with the sum of its twelve months, and month-on-month growth rates were screened for recording errors. No observations were made, and all values are expressed in nominal USD, the currency in which the firm sets its budget.

2.3. Forecasting models

Ten models in three groups were estimated, ordered from the simplest to the most structure, as can be seen on Table 3. In line with the benchmark-first principle [7], [8], each group must outperform the previous one to justify its additional complexity. Throughout, y_t denotes revenue in month t and ŷ_t its forecast.

Table 3. Model specifications

Group	Model	Specification	Horizon	Parameters
A Baseline	A1 Naïve	Last observed month	One-step, rolling	None
A Baseline	A2 Historical growth	Same month last year × (1 + training growth rate)	12-step, fixed origin	g = 3.54% (FY2024→FY2025)
A Baseline	A3 Moving average	Mean of the last three months	One-step, rolling	k = 3
B Statistical	B1 Simple exponential smoothing	Weighted average of last actual and last forecast	One-step, rolling	α = 0.30
B Statistical	B2 ARIMA(1,0,0)	First-order autoregression with intercept	One-step, rolling	c, φ estimated on training data
B Statistical	B3 Seasonal naïve	Same month last year	12-step, fixed origin	None
C Decomposition	C1 Prophet-type additive	Linear trend + month effects	12-step, fixed origin	a, b, 12-month effects
C Decomposition	C2 Holt-Winters additive	Level, trend and additive seasonality	12-step, fixed origin	α, β, γ; m = 12
C Decomposition	C3 EWT-Prophet hybrid	Smoothed component + linear trend + month effects	12-step, fixed origin	3-month window; a, b
C Decomposition	C4 EWT-Holt-Winters hybrid	Holt-Winters on the smoothed component + residual month effects	12-step, fixed origin	3-month window; α, β, γ

2.3.1. Model A: baseline methods

The baseline group requires no estimated parameters except a growth rate. The naïve model repeats the last observed month, the historical-growth model scales the same month of the previous year, and the moving average takes the meaning of the last three months. Their forecasts are defined as Equation (1), (2) and (3), where g is the annual revenue growth observed within the training period. In the back-test, g equals the FY2024 to FY2025 growth rate of 3.54%. The same rate is applied to every month of the forecast year.

$$\hat{y}_t^{A1} = y_{t-1} \quad (1)$$

$$\hat{y}_t^{A2} = y_{t-12}(1+g) \quad (2)$$

$$\hat{y}_t^{A3} = \frac{1}{3}\sum_{j=1}^{3} y_{t-j} \quad (3)$$

2.3.2. Model B: statistical time-series methods

Simple exponential smoothing updates the forecast as a weighted average of the latest observation and the previous forecast, as can be seen on Equation (4), (5), and (6) [17]. The smoothing constant was fixed at α = 0.30, a common value for monthly business data, and the level was initialized at the first observation. The ARIMA(1,0,0) model captures first-order autocorrelation, with c and φ estimated by conditional least squares on the training months only, while the seasonal naïve model repeats last year's value for the same month [18].

$$\hat{y}_{t+1}^{B1} = \alpha y_t + (1-\alpha)\hat{y}_t \quad (4)$$

$$y_t = c + \phi y_{t-1} + \varepsilon_t \quad (5)$$

$$\hat{y}_t^{B3} = y_{t-12} \quad (6)$$

A full SARIMA model was not estimated in this study. Two years of training data contain only two seasonal cycles, which are too few to estimate seasonal parameters reliably. The seasonal naïve model therefore serves as the seasonal benchmark instead.

2.3.3. Model C: decomposition-based methods

The Prophet-type model (C1) follows the additive structure of Taylor and Letham, which can be seen on Equation (7), combining a linear trend fitted to the 24 training months with a fixed effect for each calendar month [19]. Each month effect is estimated as the mean residual of that month across FY2024 and FY2025. The Holt-Winters model (C2) uses additive trend and seasonality with a 12-month period, which can be seen on Equation (8) and its parameters α, β and γ are chosen to minimize the in-sample one-step squared error [20].

$$C1: y_t = a + bt + s_{m(t)} + \varepsilon_t \quad (7)$$

$$C2: \ell_t = \alpha(y_t - s_{t-12}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \quad (8)$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \quad (9)$$

$$s_t = \gamma(y_t - \ell_t) + (1-\gamma)s_{t-12} \quad (10)$$

$$\hat{y}_{t+h} = \ell_t + hb_t + s_{t+h-12} \quad (11)$$

The hybrid models C3 and C4 follow the decomposition logic of the empirical wavelet transform [21]. The series is split into a low-frequency component, extracted with

a centered three-month moving average, and a high-frequency remainder. The low-frequency component is extrapolated by the linear trend of C1 (for C3) or by Holt-Winters (for C4), and the mean monthly remainder is then added back.

This smoothing filter approximates, rather than reproduces, the adaptive spectral segmentation of the full transform. The hybrids are therefore reported as EWT-type models. They are re-estimated with the full transform as a robustness check.

2.3.4. Forecast horizons

The one-step models (A1, A3, B1 and B2) use each FY2026 actual as soon as it becomes available, as a firm would when updating its forecast monthly. The fixed-origin models (A2, B3 and C1 to C4) forecast all twelve months of FY2026 from the end of FY2025, which mirrors the setting of an annual budget. Because one-step forecasts draw on more recent information, the results are compared within each horizon class as well as across all models.

2.4. Evaluation design

All models are evaluated on the twelve months of FY2026, which none of them used during estimation [22]. This design avoids the optimistic bias of in-sample accuracy. With e_t = y_t − ŷ_t and n = 12, accuracy and bias are measured as Equation (12), (13), (14) and (15).

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} |e_t| \quad (12)$$

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n} \left|\frac{e_t}{y_t}\right| \quad (13)$$

$$\text{Bias\%} = \frac{\sum_t \hat{y}_t - \sum_t y_t}{\sum_t y_t} \times 100 \quad (14)$$

$$\text{F/A ratio} = \frac{\sum_t \hat{y}_t}{\sum_t y_t} \quad (15)$$

MAE expresses error in USD, MAPE allows comparison with other studies, and bias reveals whether a model systematically under- or over-forecasts annual revenue [23], [24]. Bias matters because an annual budget is judged on its total, in which monthly errors of the same sign accumulate. Since revenue never approaches zero in this series, MAPE is well defined, and its asymmetry is offset by reporting MAE alongside it [25].

2.4.1. Model selection

The selected model must meet three criteria for the hold-out: the lowest MAPE, the lowest MAE and an absolute bias below 10% of annual revenue. If no model meets all three, models are ranked first by MAPE and then by MAE. A more complex model is preferred only if it also beats the naïve benchmark (A1).

2.4.2. Significance testing

Differences in accuracy between the best model and every other model are tested with the Diebold-Mariano test on absolute errors. The test uses the small-sample correction of certain study [26]. Its statistics follow a t-distribution with n − 1 = 11 degrees of freedom, which can be seen on Equation (16) and (17).

$$DM^{*} = \sqrt{\frac{n+1-2h+h(h-1)/n}{n}}\,\frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}} \quad (16)$$

$$d_t = |e_{1,t}| - |e_{2,t}| \quad (17)$$

where n = 12 and h = 1 for all comparisons. The test compares month-by-month absolute errors, and for a 12-step horizon with n = 12 the correction term n + 1 − 2h + h(h − 1)/n equals zero, so the statistic is undefined. Tests involving fixed-origin models therefore use h = 1, which understates the autocorrelation of multi-step errors, so their p-values are read conservatively at the 5% significance level.

2.4.3. Forecast combination

An equal-weight combination of the three most accurate models is also evaluated. Simple averages are robust and frequently outperform the best single model [15], [16]. The weights are not optimized, because twelve test observations are too few to estimate them reliably [27].

2.5. FY2027 forecast and prediction intervals

The best single model and the equal-weight combination selected in Section 2.4 are re-estimated on all 36 months from June 2023 to May 2026. Each model then forecasts the twelve months of FY2027 from a fixed origin in May 2026, because the budget is set once for the whole year. One-step models are iterated on their own forecasts beyond the first month, and the annual forecast is the sum of the monthly forecasts, as follows on Equation (18).

$$\hat{Y}_{\text{FY27}} = \sum_{h=1}^{12} \hat{y}_{T+h} \quad (18)$$

2.5.1. Prediction intervals

Uncertainty in the annual total is estimated by bootstrapping each model's percentage errors from the hold-out year. In each of B = 10,000 replications, twelve errors are drawn with replacement, applied to the monthly forecasts and summed into an annual total. The 2.5th, 10th, 90th and 97.5th percentiles of the simulated totals give the 80% and 95% prediction intervals, and analytical state-space intervals are computed as a cross-check for the exponential smoothing and ARIMA models [28], [29].

The bootstrap is preferred because it requires no normality assumption and reflects the errors the models actually made on FY2026 data. As most FY2026 errors were negative, resampling them carries the historical under-forecasting into the interval and gives the target the benefit of the doubt. Because months are resampled independently, however, these intervals understate the uncertainty of an annual total, so a best-case driver scenario is added as a stress test.

Each replication converts a forecast into a simulated outcome using the percentage error definition of Section 2.4, e = (y − ŷ)/y. An error drawn therefore scales the forecast upward when the model under-forecast in the hold-out year. The conversion and the annual total are defined as follows:

$$y_{T+h}^{(b)} = \frac{\hat{y}_{T+h}}{1-e_h^{(b)}} \quad (19)$$

$$Y_{\text{FY27}}^{(b)} = \sum_{h=1}^{12} y_{T+h}^{(b)}, \quad b = 1, \dots, B \quad (20)$$

2.5.2. Error variants

Three variants are simulated to test how sensitive the results are to the error assumption. The first uses the selected model with its own one-step hold-out errors, and the second uses the equal-weight combination with its own one-step errors. The third applies the hold-out errors of the most accurate fixed-origin model (C3) to the combination, which reflects the larger errors of an annual forecast and, because that model under-forecast FY2026 by about 20%, gives the most credit to FY2026's momentum.

2.5.3. Growth scenarios

Three reference scenarios frame the model forecasts. A momentum scenario repeats FY2026's growth rate of 28.8% (USD 14.17 million), and a target scenario applies the 100% growth required by the budget (USD 21.99 million). A best-case driver scenario assumes that chargeable weight grows again by 82.7% at constant FY2026 yield (USD 20.09 million).

2.6. Feasibility analysis

Feasibility is measured as the distance between the target and the FY2027 forecast distribution. The target equals T = 2 × Y_FY26, or USD 21.99 million. Four indicators are computed, as can be seen on Equation (21), (22), (23) and (24).

$$\text{Gap} = T - \hat{Y}_{\text{FY27}} \quad (21)$$

$$\text{Gap\%} = \frac{T-\hat{Y}_{\text{FY27}}}{\hat{Y}_{\text{FY27}}} \times 100 \quad (22)$$

$$\text{Coverage} = \frac{\hat{Y}_{\text{FY27}}}{T} \quad (23)$$

$$P_{\text{attain}} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left(Y_{\text{FY27}}^{(b)} \geq T\right) \quad (24)$$

The gap expresses the shortfall in USD, and coverage shows how much of the target the forecast supports. P_attain is the share of bootstrap replications in which simulated FY2027 revenue reaches the target. Together, these indicators describe both the size of the shortfall and its likelihood.

2.6.1. Decision rule

The target is classified according to Table 4. The thresholds draw on budgeting research showing that well-functioning targets are met in roughly 80 to 90% of cases [10], [30]. A target below the lowest threshold is treated as infeasible.

Table 4. Decision rule for classifying target feasibility

P_attain	Classification	Budget implication
≥ 0.80	Achievable	Retain the target
0.20 to < 0.80	Stretch	Retain only with revision clauses and funded growth initiatives [13]
< 0.20	Infeasible	Revise the target

2.6.2. Evidence-based budget

When the target is not achievable, two alternative budget levels are derived from the same distribution. The committed budget is the 20th percentile of simulated FY2027 revenue, which the firm reaches with 80% probability. The stretch budget is the median, which it reaches with 50% probability.

2.6.3. Driver check

Revenue equals chargeable weight multiplied by yield. The target can therefore be translated into the operational volume it implies. This translation is defined as Equation (25) and (26).

$$W_{\text{FY27}}^{\text{req}} = \frac{T}{\text{yield}_{\text{FY27}}} \quad (25)$$

$$\text{Required weight growth} = \frac{W_{\text{FY27}}^{\text{req}}}{W_{\text{FY26}}} - 1 \quad (26)$$

Required weight is computed under two yield assumptions: yield held at its FY2026 level, and yield continuing to change at its FY2025 to FY2026 rate. The same calculation is repeated for consignments and active traders, and the implied growth is compared with the drivers' historical growth to judge operational plausibility independently of the models. The required monthly run-rate (T ÷ 12) is also compared with the highest monthly revenue recorded in FY2026.

2.7. Validity, reliability and limitations

2.7.1. Internal validity

All parameters, including the growth rate of A2, the month effects of C1 and the coefficients of B2, are estimated from FY2024 and FY2025 only. No information from the hold-out year therefore enters model estimation. One-step models use FY2026 actuals only as lagged inputs, never as targets of their own estimation.

2.7.2. Reliability

All calculations were performed in spreadsheet workbooks with visible formulas, one workbook per model group, so every forecast can be traced to its inputs. To confirm that the results do not depend on the spreadsheets, all models were re-implemented independently in Python (NumPy and SciPy). The EWT-type hybrids were also re-estimated with a full empirical wavelet transform using a Meyer filter bank with spectrum-based boundaries [21].

2.7.3. Robustness checks

Four checks test whether the conclusions depend on design choices. The first varies the smoothing constant of B1 from $\alpha = 0.1$ to 0.9, and the second applies rolling-origin evaluation, re-estimating every model monthly and evaluating it one step ahead over the last six months of FY2026 [22]. The third re-estimates C3 and C4 with a full empirical wavelet transform [21], and the fourth replaces the bootstrap with normal-theory intervals for P_attain.

2.7.4. Limitations

The series contains only 36 monthly observations, which limits the estimation of seasonal and nonlinear models and widens prediction intervals. FY2026 also contains a structural acceleration in volume and a fall in yield, which models trained on FY2024 and FY2025 cannot anticipate. In addition, revenue is measured in nominal USD, the Prophet-type model omits Prophet's automatic changepoints and Bayesian estimation, and the single-case design limits statistical generalization, although the procedure itself is transferable to other carriers.

2.7.5. Ethics

The firm agreed to the use of its internal data on condition of anonymity. Only aggregated firm-level figures are reported in this article. No customer-level data is disclosed.

3. Result and Discussion

3.1. Descriptive analysis

Revenue was flat for two years and then rose sharply. FY2025 grew by only 3.5% over FY2024, whereas FY2026 grew by 28.8% to USD 11.00 million. Ten of the twelve months in FY2026 were higher than the best months of the two previous years, which can be seen on Figure 2. The FY2026 increase was driven by volume and customers rather than price, which can be seen on Table 5. Chargeable weight rose by 82.7% and average monthly active traders by 50.9%, while yield fell from USD 10.17 to USD 7.17 per kg (−29.5%) and revenue per consignment by 14.2%. Monthly revenue also became less volatile, with the coefficient of variation falling from 12.9% in FY2025 to 8.3% in FY2026.

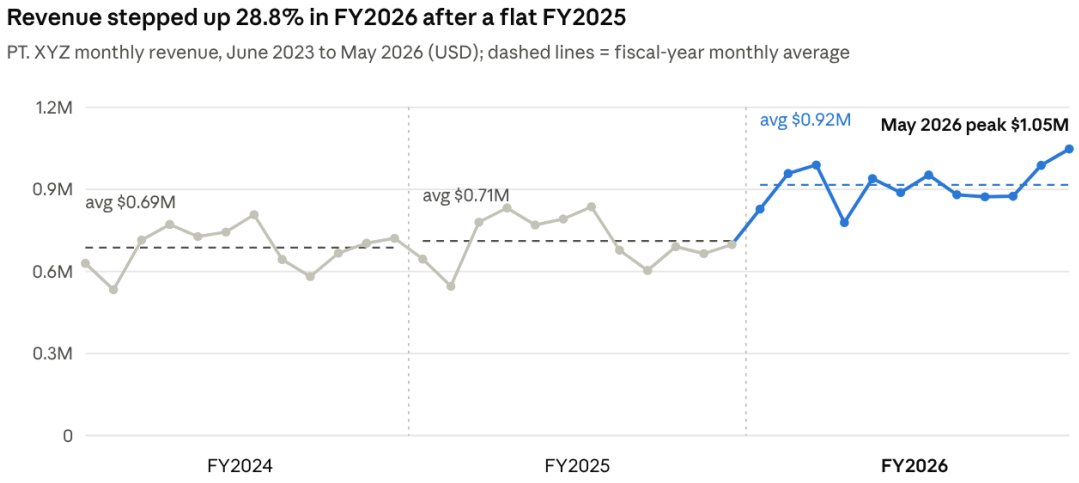


Figure 2. Monthly revenue of PT XYZ, June 2023–May 2026 (USD). Source: PT XYZ internal sales performance report

Table 5. Annual revenue and operational drivers

Fiscal year	Revenue (USD m)	Weight (kg)	Consignments (AWB)	Avg. monthly active traders	Yield (USD/kg)	Revenue per AWB (USD)	Monthly CV (%)
FY2024	8.24	756,810	57,226	515	10.89	144.04	11.5
FY2025	8.53	839,245	62,140	602	10.17	137.34	12.9
FY2026	11.00	1,533,505	93,342	908	7.17	117.80	8.3

This pattern has direct consequences for model selection. The FY2026 step change lies outside anything observed in the training years. Models that extrapolate the FY2024 to FY2025 trend and seasonality are therefore expected to under-forecast the hold-out year.

3.2. Hold-out accuracy and model selection (RQ1)

The naïve model was the most accurate on the FY2026 hold-out, with a MAPE of 9.62% and the smallest bias (−3.18%). Simple exponential smoothing (9.99%) and the three-month moving average (10.53%) were statistically indistinguishable from it. Every other model was significantly less accurate at the 5% level, which can be seen on Table 6.

Table 6. Hold-out accuracy on FY2026 (actual revenue USD 11.00 million), ranked by MAPE

Model	Horizon	MAE (USD)	MAPE (%)	Bias (%)	FY2026 forecast (USD m)	DM* vs. A1 (p-value)
A1 Naïve	One-step	85,920	9.62	−3.18	10.65	—
B1 Simple exponential smoothing	One-step	93,393	9.99	−7.93	10.12	−0.34 (0.740)
A3 3-month moving average	One-step	96,816	10.53	−5.69	10.37	−0.53 (0.606)
B2 ARIMA(1,0,0)	One-step	153,206	16.29	−16.29	9.20	−2.39 (0.036)
C3 EWT-Prophet hybrid	12-step	182,659	19.42	−19.93	8.80	−2.85 (0.016)
C1 Prophet-type additive	12-step	193,595	20.76	−20.10	8.79	−3.04 (0.011)
A2 Historical growth	12-step	193,744	20.80	−19.63	8.84	−3.01 (0.012)
B3 Seasonal naïve	12-step	214,017	22.94	−22.38	8.53	−3.35 (0.007)
C2 Holt-Winters additive	12-step	240,453	25.82	−26.24	8.11	−3.90 (0.002)
C4 EWT-Holt-Winters hybrid	12-step	314,252	33.78	−34.30	7.22	−5.10 (<0.001)
Combination (A1, B1, A3)	One-step	90,643	9.89	−5.60	10.38	−0.34 (0.743)

Note. DM* is the Harvey et al. [26] corrected Diebold-Mariano statistic on absolute errors (t with 11 df); a negative value means the model is less accurate than A1. Values in parentheses are two-sided p-values.

a. All models under-forecast. Every model produced a negative bias, ranging from −3.2% for the naïve model to −34.3% for the EWT-Holt-Winters hybrid. This confirms that FY2026 broke with the training pattern. No model estimated on FY2024 and FY2025 anticipated the step change in volume.

b. Horizon drives accuracy more than model complexity. All four one-step models outperformed all six fixed-origin models, because they absorb each new FY2026 actual as it arrives. Within the fixed-origin class, which corresponds to setting an annual budget, the EWT-Prophet hybrid was the most accurate (19.42%). However, it did not differ significantly from the historical-growth model ($p = 0.567$), the Prophet-type model ($p = 0.494$) or the seasonal naïve benchmark ($p = 0.098$).

c. Estimated parameters. Re-estimated on the 24 training months, the AR(1) coefficient was $\varphi = 0.345$ ($t = 1.72$, $R^2 = 0.12$), which is not significant at the 5% level. This weak persistence explains why ARIMA(1,0,0) pulled its forecasts back toward the training mean and under-forecast by 16.3%. Holt-Winters selected $\alpha = 0.85$ with trend and seasonal smoothing close to zero, which effectively reduced it to a level model with fixed seasonal effects.

d. Combination. The equal-weight combination of the three best models reached a MAPE of 9.89%. This was not better than the naïve model alone. The difference was also not statistically significant ($p = 0.743$).

e. Selected model. The naïve model meets all three selection criteria: the lowest MAPE, the lowest MAE and an absolute bias below 10%. It is therefore used as the primary model for FY2027. The combination is reported alongside it as a sensitivity check.

3.3. Robustness checks

The robustness checks confirm the ranking. Models that track the most recent level perform best, and models that extrapolate the training pattern perform worst. The three checks reported in this section are discussed in turn.

a. Smoothing constant. Hold-out MAPE for simple exponential smoothing fell steadily as α increased, from 14.55% at $\alpha = 0.1$ to 9.99% at $\alpha = 0.3$ and 9.57% at $\alpha = 0.9$. The optimum therefore lies at the naïve end of the range ($\alpha \to 1$). The advantage of the naïve model is thus not an artefact of the fixed $\alpha = 0.30$.

b. Rolling origin. When every model is re-estimated each month and evaluated one step ahead over the last six months of FY2026 (December 2025 to May 2026), the naïve model again ranks first, as can be seen on Table 7. Once they are allowed to learn from FY2026 data, the decomposition models close most of their gap. This confirms that their hold-out weakness came from the structural break rather than from their specification.

c. Full empirical wavelet transforms. Re-estimating the hybrids with a full EWT (Meyer filter bank with boundaries at 0.26 and 0.79 rad, giving three modes) produced hold-out MAPEs of 20.67% for the trend-based variant and 23.21% for the Holt-Winters-based variant. These values are close to the EWT-type results (19.42% and 33.78%), and both variants remain significantly less accurate than the naïve model ($p = 0.012$ and $p = 0.006$). The moving-average approximation therefore does not change the conclusions.

Table 7. Rolling-origin one-step MAPE, December 2025–May 2026

Model	MAPE (%)
A1 Naïve	5.52
C4 EWT-Holt-Winters hybrid	6.03
B1 Simple exponential smoothing	6.17
C2 Holt-Winters additive	6.62
C3 EWT-Prophet hybrid	6.68
A3 3-month moving average	7.56
B2 ARIMA(1,0,0)	8.04
C1 Prophet-type additive	10.88
A2 Historical growth	12.81
B3 Seasonal naïve	25.53

3.4. FY2027 revenue forecast (RQ2)

Re-estimated on all 36 months, the selected naïve model forecasts FY2027 revenue of USD 12.57 million, or 14.3% above FY2026, holding every month at the May 2026 level of USD 1.05 million. The equal-weight

combination forecasts USD 11.99 million (+9.1%). Across all ten models, FY2027 forecasts range from USD 10.34 million to USD 14.35 million, or from −5.9% to +30.5% growth, as can be seen on Table 8.

Table 8. FY2027 point forecasts, all models re-estimated on FY2024–FY2026

Model	FY2027 forecast (USD m)	Growth vs. FY2026 (%)
C2 Holt-Winters additive	14.35	+30.5
A2 Historical growth	14.17	+28.8
A1 Naïve (selected)	12.57	+14.3
C4 EWT-Holt-Winters hybrid	12.38	+12.6
Combination (A1, B1, A3)	11.99	+9.1
A3 3-month moving average	11.96	+8.8
C1 Prophet-type additive	11.82	+7.5
C3 EWT-Prophet hybrid	11.81	+7.4
B1 Simple exponential smoothing	11.45	+4.1
B3 Seasonal naïve	11.00	0.0
B2 ARIMA(1,0,0)	10.34	−5.9

The highest forecasts come from models that carry FY2026's growth forward. Holt-Winters does so through its updated level, and the historical-growth model does so by construction because its growth rate is now the FY2025 to FY2026 rate. No model forecasts growth above 31%.

Bootstrapping the naïve model's hold-out errors gives a bias-adjusted median of USD 13.10 million, with an 80% interval of USD 12.54 to 13.62 million and a 95% interval of USD 12.23 to 13.88 million. The median lies above the point forecast because the model under-forecast FY2026 by 3.2%, and the bootstrap carries this correction forward.

These intervals apply one-step errors to a 12-month horizon and resample months independently. They therefore understate the uncertainty of an annual total. The 12-step error variant in Table 9 and Figure 3 provides a wider and more conservative range, and Figure 3 also shows the intervals for the combination against the target.

3.5. Feasibility of the 100% target (RQ3)

The 100% incremental target is infeasible on every measure. The selected forecast covers only 57% of the target, and none of the 10,000 bootstrap replications reaches it. Even a best-case scenario that repeats FY2026's 82.7% volume growth at constant yield falls USD 1.90 million short, which can be seen on Figure 3.

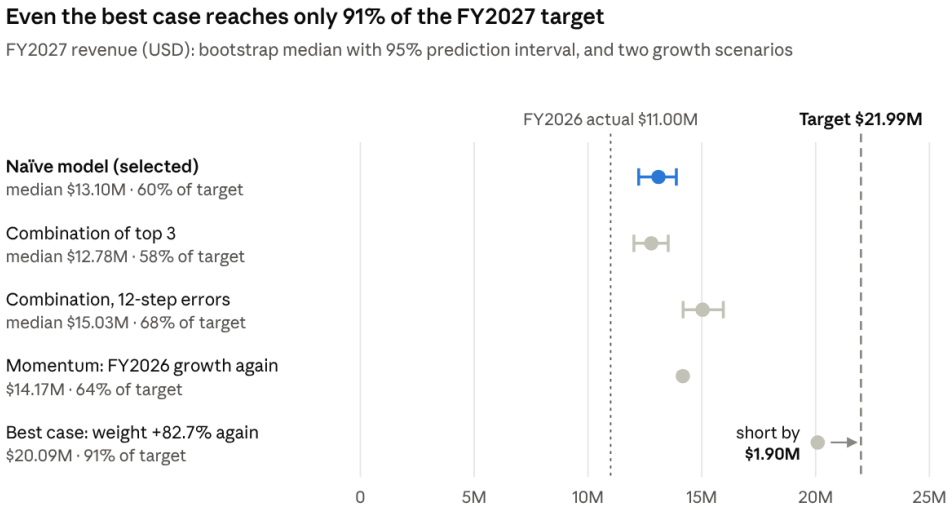


Figure 3. FY2027 revenue forecasts and growth scenarios against the USD 21.99 million target (bootstrap medians with 95% prediction intervals). Source: authors' calculations from PT XYZ data; 10,000 bootstrap replications per variant.

Table 9. Feasibility indicators for the USD 21.99 million target

Indicator	Naïve (selected)	Combination	Combination, 12-step errors
Point forecast (USD m)	12.57	11.99	11.99
Gap to target (USD m)	9.42	10.00	10.00
Gap (% of forecast)	75.0%	83.4%	83.4%
Coverage (forecast ÷ target)	0.57	0.55	0.55
P_attain, bootstrap	0.000	0.000	0.000
P_attain, normal theory	<0.001	<0.001	<0.001
80% prediction interval (USD m)	12.54–13.62	12.27–13.27	14.48–15.63
Committed budget, P20 (USD m)	12.74	12.44	14.67
Highest simulated outcome (USD m)	14.58	14.21	17.18

a. Classification. With P_attain below 0.20 under every error assumption, the target falls into the infeasible category of the decision rule (Section 2.6). The result does not depend on the bootstrap, since normal-theory intervals lead to the same conclusion. Even the highest of 30,000 simulated outcomes across the three variants (USD 17.18 million) remains USD 4.81 million below the target.

b. Evidence-based budget. The selected model supports a committed budget of USD 12.74 million (+15.8% on FY2026), reached with 80% probability, and a stretch budget of USD 13.10 million (+19.1%), reached with 50% probability. Under the wider 12-step error assumption, which gives the most credit to FY2026's momentum, the two budgets rise to USD 14.67 million (+33.4%) and USD 15.03 million (+36.7%). Even these upper values remain well below the USD 21.99 million target.

c. Driver check. Translating the target into operations shows why it is out of reach, which can be seen on Table 10. At FY2026 unit values, every driver would need to grow by 100%. If unit values keep falling at their FY2025 to FY2026 rate, weight would need to grow by 184%, more than twice its record growth of 82.7%.

Table 10. Operational growth required by the target, compared with history

Driver	Growth FY2024→FY2025	Growth FY2025→FY2026	Required, unit value flat	Required, unit value keeps falling
Weight (kg), unit value = yield per kg	+10.9%	+82.7%	+100.0%	+183.7%
Consignments (AWB), unit value = revenue per AWB	+8.6%	+50.2%	+100.0%	+133.2%
Active traders, unit value = revenue per trader-month	+16.9%	+50.9%	+100.0%	+134.2%

d. Run-rate. Reaching USD 21.99 million requires average monthly revenue of USD 1.83 million. This is 1.75 times the best month PT XYZ has ever recorded (USD 1.05 million in May 2026) and twice the FY2026 monthly average. Such a level would have to be sustained for twelve consecutive months.

3.6. Summary of propositions

Two of the three propositions are supported, as can be seen on Table 11. P1 and P3 are confirmed by the hold-out and feasibility results. P2 is not supported, for the reasons explained on Table 11.

Table 11. Outcome of the research propositions

Proposition	Outcome	Evidence
P1. Simple statistical models forecast at least as accurately as decomposition-based models	Supported	Simple exponential smoothing (9.99%) and ARIMA(1,0,0) (16.29%) were more accurate than every decomposition model (19.42–33.78%). Within the 12-step class, no decomposition model was significantly more accurate than the seasonal naïve benchmark (best p = 0.098).
P2. An equal-weight combination beats the single best model	Not supported	The combination's MAPE of 9.89% was higher than the naïve model's 9.62% (p = 0.743).
P3. The target lies above the upper bound of the forecast's prediction interval	Supported	The target exceeds the upper 95% bound in all three variants (USD 13.52–15.95 million); P_attain = 0.000.

The failure of P2 is consistent with the structural break in FY2026. Forecast combination works when model errors differ, but here all three components under-forecast for the same reason. Averaging the naïve model with smoothers that lag the new level therefore dilutes, rather than diversify, its accuracy. The results answer the central question clearly: the 100% incremental target is not supported by any model, error assumption or operational scenario. Sections 3.7 to 3.9 interpret these findings against the literature. They also set out their theoretical and managerial implications.

3.7. Interpretation of the findings

3.7.1. Simple models and the structural break

The superiority of the naïve model is consistent with large-scale evidence that simple benchmarks are hard to beat, particularly on short series [7], [8], [14]. The mechanism, however, is specific to this case, because FY2026 marked a level shift driven by volume and customer growth. Models that extrapolated the FY2024 to FY2025 pattern under-forecast this shift by 16% to 34%, whereas models that track the most recent level absorbed it fastest.

The rolling-origin check shows that the decomposition models were not mis specified as such. Once they could learn from FY2026 data, their one-step accuracy approached that of the naïve model. Their hold-out weakness therefore reflects the break rather than the method, echoing warnings that complex models need long and stable histories [31].

3.7.2. Horizon versus complexity

The clearest pattern in Table 6 is that forecast horizon explains accuracy better than model complexity, since every one-step model beat every fixed-origin model. This matters for budgeting, because an annual budget is

a fixed-origin forecast by construction. The accuracy a firm can expect from its budget is therefore closer to the 19% to 34% MAPE of the fixed-origin models than to the 10% of the one-step models, a distinction that model comparisons in logistics rarely make [32], [33].

3.7.3. The combination puzzle under a break

Forecast combination did not improve on the best single model, contrary to the general finding [15], [16]. Combination works when errors are diverse, but here all three components under-forecast for the same reason, so averaging diluted the accuracy of the naïve model. This outcome is consistent with theoretical accounts of when equal weights lose their advantage [27], [34].

3.7.4. Why the target fails

The models and the operational drivers agree on why the target is infeasible. The most generous forecast, which carries FY2026's momentum forward, reaches about USD 15 million, and no simulated outcome exceeded USD 17.18 million. Independently of any model, the target would require chargeable weight to grow by 100% at constant yield, or by 184% if yield continued to fall.

FY2026 also shows that volume growth was partly bought with lower yield. Volume-led expansion therefore cannot be assumed to translate into revenue one for one. Any plan that relies on volume alone risks repeating the yield dilution observed in FY2026.

3.8. Theoretical implications

The study makes three contributions to theory. First, it operationalizes target difficulty ex ante, whereas budgeting research typically measures difficulty after the fact as the rate of target achievement [10], [30]. Expressing difficulty as the probability of attainment under a validated forecast allows it to be assessed when the target is set, which is when it can still be changed.

Second, the findings refine the application of goal-setting theory to budgets. A study argue that difficult goals motivate only while they remain attainable, so a target with a probability of attainment near zero is expected to weaken commitment rather than raise effort [11]. The PT XYZ target, set after a year of 28.8% growth, also fits the ratcheting pattern described by multiple studies [35], [36], with the risk of effort withdrawal shown on different research [12].

Third, the study connects two literatures that seldom meet. Forecasting research separates forecasts from targets and plans [8], while budgeting research documents the tension between the planning and motivational roles of the budget [37], [38]. The feasibility procedure turns this conceptual separation into a measurable quantity: the distance between what management wants and what the evidence supports.

3.9. Managerial implications

The results point to four actions for PT XYZ. Each action addresses one source of the gap between the target and the forecast. The actions are described in turn below.

a. Separate the forecast, the target and the plan. The FY2027 operating budget should be set at the committed level of about USD 12.7 million, which the firm reaches with 80% probability, with a stretch target of about USD 13.1 million. If management keeps USD 21.99 million, it should be considered an aspirational goal. It should not be used for planning related to capacity, staffing, or cash flow.

b. Fund the gap explicitly. Any revenue above the forecast must come from identified initiatives, such as new trade lanes, key-account acquisition or pricing actions. Each initiative should have its own quantified contribution and owner. Because difficult targets help only when they can be revised [13], the budget should also include a mid-year revision clause.

c. Manage yield as well as volume. Revenue per kilogram fell by 29.5% in FY2026. Sales incentives tied only to volume risk repeating this dilution. Incentives should therefore reward revenue or contribution margin rather than shipped weight alone.

d. Forecast monthly budget annually. One-step forecasts were far more accurate than fixed-origin ones in this study. The firm should therefore complement the annual budget with a rolling monthly forecast. This practice is consistent with the improved budgeting approaches reported by studies before [39], [40].

For other logistics carriers, the procedure is readily transferable. It requires only a monthly revenue series, spreadsheet or open-source software, and a target to test. This makes it suitable for small and medium firms in emerging markets, where forecasting resources are limited [2].

4. Conclusion

This study tested whether PT XYZ's FY2027 sales budget, which require revenue to grow by 100% to USD 21.99 million, is quantitatively feasible by comparing ten forecasting models on 36 months of data. The naïve model forecast FY2026 revenue most accurately (MAPE 9.62%), and no decomposition-based model beat the simple benchmarks, showing that forecast horizon mattered more than model complexity (RQ1). FY2027 revenue is forecast at USD 12.57 million (+14.3%), with plausible outcomes across all models and error assumptions between about USD 10 and 17 million (RQ2). The target exceeds the selected forecast

by USD 9.42 million and even the most generous estimate (USD 15.03 million) by USD 6.96 million, its probability of attainment is effectively zero, and it would require volumes to grow at more than twice their record rate; it is therefore infeasible, and a committed budget of USD 12.74 million with a stretch budget of USD 13.10 million is proposed instead (RQ3). These conclusions are limited by the short series, intervals that understate annual uncertainty and the single-case design, so future research should use longer histories or panels of carriers, driver-based models of volume and yield, and field evidence on how sales teams respond to low-probability targets.

References

[1] Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science, 43*(4), 546–558. https://doi.org/10.1287/mnsc.43.4.546

[2] Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research, 252*(1), 1–26. https://doi.org/10.1016/j.ejor.2015.11.010

[3] Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transportation Research Part E: Logistics and Transportation Review, 136*. https://doi.org/10.1016/j.tre.2020.101922

[4] Notteboom, T., Pallis, T., & Rodrigue, J.-P. (2021). Disruptions and resilience in global container shipping and ports: The COVID-19 pandemic versus the 2008–2009 financial crisis. *Maritime Economics & Logistics, 23*(2), 179–210. https://doi.org/10.1057/s41278-020-00180-5

[5] UPS. (2026). *UPS releases 4Q 2025 earnings and provides 2026 guidance*. Business Wire. Retrieved from https://www.businesswire.com/news/home/20260127727273/en/UPS-Releases-4Q-2025-Earnings-and-Provides-2026-Guidance

[6] Anugranov, D. (2026). *Indonesia pushes multimodal logistics to cut costs at 14.29 percent GDP*. Radio Republik Indonesia. Retrieved from https://rri.co.id/en/national/2379761/indonesia-pushes-multimodal-logistics-to-cut-costs-at-1429-percent-gdp

[7] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting, 36*(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

[8] Petropoulos, F., Apiletti, D., Assimakopoulos, V., et al. (2022). Forecasting: Theory and practice. *International Journal of Forecasting, 38*(3), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

[9] Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application, 1*, 125–151. https://doi.org/10.1146/annurev-statistics-062713-085831

[10] Merchant, K. A., & Manzoni, J.-F. (1989). The achievability of budget targets in profit centers: A field study. *The Accounting Review, 64*(3), 539–558.

[11] Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

[12] Bouwens, J., & Kroos, P. (2011). Target ratcheting and effort reduction. *Journal of Accounting and Economics, 51*(1–2), 171–185. https://doi.org/10.1016/j.jacceco.2010.07.002

[13] Arnold, M. C., & Artz, M. (2015). Target difficulty, target flexibility, and firm performance: Evidence from business units' targets. *Accounting, Organizations and Society, 40*, 61–77. https://doi.org/10.1016/j.aos.2014.12.002

[14] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS ONE, 13*(3). https://doi.org/10.1371/journal.pone.0194889

[15] Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly, 20*(4), 451–468. https://doi.org/10.1057/jors.1969.103

[16] 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. https://doi.org/10.1016/j.ijforecast.2022.11.005

[17] Gardner, E. S., Jr. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting, 22*(4), 637–666. https://doi.org/10.1016/j.ijforecast.2006.03.005

[18] Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3), 1–22. https://doi.org/10.18637/jss.v027.i03

[19] Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician, 72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

[20] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science, 6*(3), 324–342. https://doi.org/10.1287/mnsc.6.3.324

[21] Gilles, J. (2013). Empirical wavelet transform. *IEEE Transactions on Signal Processing, 61*(16), 3999–4010. https://doi.org/10.1109/TSP.2013.2265222

[22] Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting, 16*(4), 437–450. https://doi.org/10.1016/S0169-2070(00)00065-0

[23] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

[24] Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting, 29*(3), 510–522. https://doi.org/10.1016/j.ijforecast.2012.09.002

[25] Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting, 8*(1), 69–80. https://doi.org/10.1016/0169-2070(92)90008-W

[26] Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting, 13*(2), 281–291. https://doi.org/10.1016/S0169-2070(96)00719-4

[27] Smith, J., & Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics, 71*(3), 331–355. https://doi.org/10.1111/j.1468-0084.2008.00541.x

[28] Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting, 18*(3), 439–454. https://doi.org/10.1016/S0169-2070(01)00110-8

[29] Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business & Economic Statistics, 11*(2), 121–135. https://doi.org/10.1080/07350015.1993.10509938

[30] Indjejikian, R. J., Matějka, M., Merchant, K. A., & Van der Stede, W. A. (2014). Earnings targets and annual bonus incentives. *The Accounting Review, 89*(4), 1227–1258. https://doi.org/10.2308/accr-50732

[31] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting, 37*(1), 388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008

[32] Xie, G., Wang, S., Zhao, Y., & Lai, K. K. (2013). Hybrid approaches based on LSSVR model for container throughput forecasting: A comparative study. *Applied Soft Computing, 13*(5), 2232–2241. https://doi.org/10.1016/j.asoc.2013.02.002

[33] Liu, J., Chu, N., Wang, P., Zhou, L., & Chen, H. (2024). A novel hybrid model for freight volume prediction based on the Baidu search index and emergency. *Neural Computing and Applications, 36*(3), 1313–1328. https://doi.org/10.1007/s00521-023-09106-7

[34] Claeskens, G., Magnus, J. R., Vasnev, A. L., & Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting, 32*(3), 754–762. https://doi.org/10.1016/j.ijforecast.2015.12.005

[35] Weitzman, M. L. (1980). The "ratchet principle" and performance incentives. *The Bell Journal of Economics, 11*(1), 302–308. https://doi.org/10.2307/3003414

[36] Leone, A. J., & Rock, S. (2002). Empirical tests of budget ratcheting and its effect on managers' discretionary accrual choices. *Journal of Accounting and Economics, 33*(1), 43–67. https://doi.org/10.1016/S0165-4101(01)00044-1

[37] Hansen, S. C., Otley, D. T., & Van der Stede, W. A. (2003). Practice developments in budgeting: An overview and research perspective. *Journal of Management Accounting Research, 15*(1), 95–116. https://doi.org/10.2308/jmar.2003.15.1.95

[38] Covaleski, M. A., Evans, J. H., III, Luft, J. L., & Shields, M. D. (2003). Budgeting research: Three theoretical perspectives and criteria for selective integration. *Journal of Management Accounting Research, 15*(1), 3–49. https://doi.org/10.2308/jmar.2003.15.1.3

[39] Libby, T., & Lindsay, R. M. (2010). Beyond budgeting or budgeting reconsidered? A survey of North-American budgeting practice. *Management Accounting Research, 21*(1), 56–75. https://doi.org/10.1016/j.mar.2009.10.003

[40] Ekholm, B.-G., & Wallin, J. (2000). Is the annual budget really dead? *European Accounting Review, 9*(4), 519–539. https://doi.org/10.1080/09638180020024007