Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Forecasting University Enrolment Amid Crisis and Admission Restructuring: A Phase-Aware Regression Model For A 15-Year Series (2011–2025)

Vo Van On[1] , Bui Tat To[2] , Nguyen Thi Hong Tham[3]

[1]Institute of Pharmaceuticals and Health Foods, Thu Dau Mot University, Ho Chi Minh City, Vietnam
[2]Faculty of Economics, Binh Duong University, Ho Chi Minh City, Vietnam
[3]School of Law and Management, Thu Dau Mot University, Ho Chi Minh City, Vietnam
Corresponding authors: Vo Van On (onvv@tdmu.edu.vn); Nguyen Thi Hong Tham (thamnth@tdmu.edu.vn)

ABSTRACT
Enrolment forecasting is a core input to planning teaching staff, classrooms, laboratories, dormitories and budgets. Institutional series, however, are short and may contain regime changes caused by external shocks or by reforms of admission policy. This study uses a 15-year enrolment series (2011–2025) from Thu Dau Mot University, Ho Chi Minh City, Vietnam, to test for structural breaks and to develop a Phase-Aware Weighted Least Squares (PA-WLS) model. The specification combines a time index rebased at 2011 with two phase indicators — the COVID-19 period (2020–2022) and the admission restructuring period (2023–2024) — and downweights the volatile observations from 2020–2024.
The Chow test identifies a significant break at 2023 ($F = 7.098$, $p = 0.011$), and the joint test of the phase effects is also significant ($F = 5.328$, $p = 0.024$). PA-WLS estimates a baseline trend of 175.72 students per year, a level shift of +1,903.26 for the pandemic phase, and a level shift of −1,979.21 for the restructuring phase. Across eight rolling origins, the MAPE is 16.22%, compared with 27.28% for a linear trend over time. With only eight origins, this 11.06-point gap does not reach significance at the 5% level (paired $t = -1.437$, $p = 0.194$; Diebold–Mariano DM = −1.707, $p = 0.132$). None of the four nonlinear machine-learning algorithms outperforms the phase-aware specifications under the same protocol; the tree ensembles achieve a near-perfect in-sample fit yet forecast the worst out of sample.
Re-estimation on the full series yields point forecasts of 6,390 students for 2026 and 6,566 for 2027, with 95% prediction intervals of [4,274; 8,506] and [4,368; 8,763]. Sensitivity analysis shows that the choice of specification shifts the 2026 forecast by 6,001–6,861 students, roughly 37 times the range induced by the weighting parameter. The practical value of the procedure lies less in a single point estimate than in identifying regime change, validating out-of-sample, and quantifying uncertainty for scenario-based planning.

Keywords: enrolment forecasting; structural break; weighted regression; phase dummy variables; machine learning; data-driven educational management; COVID-19.

1. INTRODUCTION

In higher education, operational data are increasingly treated as an input to strategic decisions rather than as material for administrative reporting alone. For admissions management, the size of the incoming cohort is directly tied to the availability of teaching staff, classrooms, laboratories, dormitories, budgets, and student support services. Errors in either direction carry costs: over-forecasting produces idle capacity and fixed costs, while under-forecasting places pressure on facilities, staffing, and services. A forecast with a small point error but a wide uncertainty interval therefore still demands cautious interpretation.
Institutional enrolment forecasting differs from most educational data mining problems in one decisive respect. Whereas learner-level data may contain thousands or millions of records, institution-wide data usually yield one observation per year. For a 15-year series, the sample size is 15. Algorithms with many parameters overfit readily under such conditions, and in-sample performance becomes an unreliable guide to forecasting ability.
The problem becomes harder still when the series spans regime changes. At the institution studied here, the period 2020–2022 was strongly affected by the pandemic, while 2023–2024 coincided with substantial change in admission methods following the early-admission mechanism of Circular 08/2022/TT-BGDĐT and its repeal by Circular 06/2025/TT-BGDĐT. There is therefore an institutional basis for examining the series in terms of phases rather than requiring all fifteen observations to lie on a single linear trend.

1.1. Research gap and questions

Two gaps motivate this study. First, the enrolment-forecasting literature is dominated by applicant-level or course-registration data with large numbers of records, and its methods do not transfer unchanged to a univariate series of fifteen annual totals. Second, although interrupted time series methods provide an established basis for modeling level or slope changes at specified time points, they are rarely combined with rolling-origin validation and explicit prediction intervals in an institutional planning context. The study therefore addresses a narrow but practically meaningful question: in a very short enrolment series that nevertheless contains distinct historical phases, does information about the phase yield more forecasting value than an increase in algorithmic complexity? Four research questions follow.

- RQ1. Does the 2011–2025 series provide statistical evidence of a structural break, and in which year?
- RQ2. Does the inclusion of phase dummy variables improve out-of-sample forecasts relative to a plain linear regression on time?
- RQ3. Do SVR, Random Forest, Gradient Boosting and XGBoost outperform parametric regressions on a series of fifteen observations?
- RQ4. How uncertain are the 2026–2027 forecasts, and what does this imply for planning?

1.2. Contributions

The study contributes a reproducible procedure rather than a single figure: testing for structural breaks before modeling; constructing a re-based time variable and phase indicators; validating over multiple rolling origins with re-estimation; benchmarking against both simple and machine-learning alternatives under an identical protocol; reporting prediction intervals; and quantifying sensitivity to both the weighting parameter and the choice of specification. Throughout, a single absolute percentage error is reported as APE, never as MAPE, and in-sample fit is not used as the primary criterion for model selection.

2. LITERATURE REVIEW

2.1. Enrolment forecasting and data-driven educational management

Long and Siemens [1] frame analytics in education as a mechanism for converting data into information for decision-making. Romero and Ventura [2] document the growth of educational data mining and learning analytics into a broad field in which prediction forms an important group of tasks. Aldowah, Al-Samarraie and Fauzy [3], synthesizing 402 studies published between 2000 and 2017 and classifying them into learning, predictive, behavioral and visualization analytics, show that predictive applications occupy a substantial place in higher education. Most of this evidence, however, comes from individual-level data or from datasets with large numbers of observations. At the institutional level, the managerial role of forecasting is different: it is an input to decisions with long lead times. The World Bank sector report on Vietnamese higher education [4] correspondingly emphasizes management capacity and the use of information in improving system performance.

Studies dealing directly with enrolment reflect the same data-rich orientation. Shao and colleagues [5] used classification and regression trees and Random Forest to predict course enrolment from student records at San Diego State University. Soltys and colleagues [6] built a machine-learning framework deployed on AWS SageMaker that used three years of admission history to estimate the enrolment probability of each applicant. Al-Naymat and Al-Betar [7] proposed a framework for predicting university student enrolment, along with a comparative evaluation of classification and forecasting models using historical enrolment data. These studies are of clear methodological value, but a model that performs well on thousands of records is not necessarily suited to a series of fifteen points.

2.2. Statistical and machine-learning methods in small samples

Makridakis, Spiliotis and Assimakopoulos [8] argue that the use of machine learning for time series forecasting requires empirical verification rather than the assumption that a more complex model will be more accurate. The M4 Competition, covering 100,000 series, showed that statistical methods and their combinations remain highly competitive [9]; the M5 Competition, using much larger hierarchical retail data, showed that machine-learning methods can be effective where data are abundant [10]. The relevant message is not that machine learning is inferior to statistics, but that model complexity must be commensurate with the information available.

The algorithms retained here as benchmarks are standard. Support Vector Regression controls model complexity through an epsilon-insensitive loss [11]; Random Forest is a bagged tree ensemble [12]; gradient boosting is stagewise additive modeling in function space [13]; and XGBoost is an efficient boosting implementation [14]. They are included not because they are expected to win, but to test directly whether nonlinearity yields any benefit at n = 15.

2.3. The pandemic and changes in admission regulation

The National Science Board [15] recorded a decline in postsecondary enrolment in the United States during the pandemic, but the effect was uneven across segments: relative to spring 2020, total postsecondary enrolment fell by about 3.5%, undergraduate enrolment by 4.9% and community college enrolment by 9.5%, while graduate enrolment continued to grow, rising by 4.6%. The signature of a global shock should therefore not be imposed in advance on every institution. In the present case, the institution's own data show that enrolment in fact rose over 2020–2022, and the phase variable is accordingly used as a descriptive indicator of regime rather than as causal evidence.

In the Vietnamese regulatory context, Circular 08/2022/TT-BGDĐT, dated 6 June 2022, permitted institutions to conduct early admission under Article 18 [16]. Circular 06/2025/TT-BGDĐT of 19 March 2025 repealed that article and took effect on 5 May 2025 [17]. This change provides the historical basis for treating 2023–2024 as a distinct phase and 2025 as the anchor point of a new state. After the study period, the framework continued to evolve: Circular 06/2026/TT-BGDĐT, promulgated on 15 February 2026, introduced an entirely new admission regulation under Higher Education Law No. 125/2025/QH15, replacing Circular 08/2022/TT-BGDĐT [18].

2.4. Structural breaks and forecast validation

Chow [19] provides the classical test of equality of coefficients between two regressions. In interrupted time series designs, level and slope indicators are used to quantify change at the point of intervention [20,21]. The present study uses level indicators only, because the aim is to separate the baseline of each phase while retaining a common underlying trend, and because two of the phases contain too few observations to identify separate slopes.

For validation, Tashman [22] sets out the benefits of rolling-origin evaluation with re-estimation as the origin advances. Bergmeir and Benítez [23] show that conventional cross-validation cannot be applied directly to dependent data and propose blocked schemes, while Bergmeir, Hyndman and Koo [24] establish the conditions — purely autoregressive specifications with uncorrelated errors — under which standard K-fold cross-validation remains valid. Since the specification adopted here is not autoregressive, and since the evaluation is intended to mirror operational one-step-ahead forecasting, the study follows the rolling-origin protocol rather than random train/test splitting.

3. METHODS

3.1. Data

The data comprise the annual total of officially enrolled students at Thu Dau Mot University from 2011 to 2025. The series is an institution-level aggregate compiled by academic year and supplied by the University's academic affairs unit; it contains no personal information. Because only fifteen annual totals are available, no further explanatory variables — tuition fees, admission cut-off scores, size of the applicant pool, number of applications — are included; these are not present in the dataset, and with eleven residual degrees of freedom they could not be identified even if they were.

The series is divided into four phases determined from policy history and the epidemiological record, and only then tested against the data, so that no breakpoint is selected merely to optimize fit. P1 (2011–2019) is the baseline; P2 (2020–2022) is the pandemic period; P3 (2023–2024) is the period of admission restructuring; and P4 (2025) is the new-normal year, reserved as an anchor point for validation. Table 1 reports the series together with the phase labels, the two phase indicators and the weights used in estimation.

Table 1. Enrolment 2011–2025, phase labels, phase indicators and model weights

Year	New entrants	Phase	D_C	D_R	ω	Growth (%)
2011	2,523	P1 – Baseline	0	0	1.0	—
2012	4,854	P1 – Baseline	0	0	1.0	+92.39
2013	3,835	P1 – Baseline	0	0	1.0	−20.99
2014	5,330	P1 – Baseline	0	0	1.0	+38.98
2015	5,336	P1 – Baseline	0	0	1.0	+0.11
2016	3,843	P1 – Baseline	0	0	1.0	−27.98
2017	4,405	P1 – Baseline	0	0	1.0	+14.62

Year	New entrants	Phase	D_C	D_R	ω	Growth (%)
2018	4,959	P1 – Baseline	0	0	1.0	+12.58
2019	4,956	P1 – Baseline	0	0	1.0	−0.06
2020	6,598	P2 – Pandemic	1	0	0.3	+33.13
2021	8,547	P2 – Pandemic	1	0	0.3	+29.54
2022	7,099	P2 – Pandemic	1	0	0.3	−16.94
2023	3,646	P3 – Restructuring	0	1	0.3	−48.64
2024	4,297	P3 – Restructuring	0	1	0.3	+17.86
2025	6,287	P4 – New normal	0	0	1.0	+46.31

Note. D_C and D_R are the pandemic and restructuring indicators defined in equation (1); ω is the estimation weight defined in equation (2). Growth is the year-on-year percentage change.

3.2. The phase-aware weighted model

Let i = 1, …, 15 index the observations and $\tau_i \in \{2011, \ldots, 2025\}$ denote the calendar year. The time variable is re-based at 2011, so that t_i runs from 0 to 14, and the two-phase indicators are defined through the indicator function $I(\cdot)$:

$$t_i = \tau_i - 2011, \qquad D_{C,i} = I(2020 \le \tau_i \le 2022), \qquad D_{R,i} = I(2023 \le \tau_i \le 2024) \tag{1}$$

Re-basing allows the intercept to be read directly as the expected baseline level in 2011 and avoids an intercept of the order of hundreds of thousands, which would carry no interpretable meaning. Observations in 2020–2024 receive a reduced weight in order to limit the influence of the most volatile years; the main specification uses ω = 0.3:

$$\omega_i = \begin{cases} 0.3, & 2020 \le \tau_i \le 2024 \\ 1, & \text{otherwise} \end{cases} \qquad i = 1, \ldots, 15 \tag{2}$$

The sum of the weights, $\Sigma\omega_i = 11.5$, is a purely descriptive indicator of the information remaining after down-weighting. It is used in no test. All inference rests on the actual sample size n = 15 and p = 4 parameters, that is, on n − p = 11 residual degrees of freedom. The model is

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 D_{C,i} + \beta_3 D_{R,i} + \varepsilon_i, \qquad E(\varepsilon_i) = 0, \qquad Var(\varepsilon_i) = \frac{\sigma^2}{\omega_i} \tag{3}$$

where β_0 is the baseline level in 2011, β_1 the baseline annual trend, and β_2 and β_3 the level shifts of the two irregular phases. The assumption that the error variance is inversely proportional to the weight is what makes equations (4), (5) and (6) valid. Estimation minimises the weighted residual sum of squares $Q(\beta) = \sum_{i=1}^{n} \omega_i (y_i - x_i'\beta)^2$, with $x_i = (1, t_i, D_{C,i}, D_{R,i})'$, giving

$$\hat{\beta} = (X'\Omega X)^{-1} X'\Omega y, \qquad \Omega = \mathrm{diag}(\omega_1, \ldots, \omega_n), \qquad s^2 = \frac{Q(\hat{\beta})}{n-p} \tag{4}$$

In-sample fit is described by the weighted coefficient of determination, in which $\bar{y}_w$ is the weighted mean:

$$R_w^2 = 1 - \frac{\sum_{i=1}^{n} \omega_i e_i^2}{\sum_{i=1}^{n} \omega_i (y_i - \bar{y}_w)^2}, \qquad \bar{y}_w = \frac{\sum_{i=1}^{n} \omega_i y_i}{\sum_{i=1}^{n} \omega_i} \tag{5}$$

The Breusch–Pagan test [25] is applied to the weight-standardized residuals. Its F form gives p = 0.023 and its LM form p = 0.069, so the evidence of heteroscedasticity is borderline but sufficient to justify reporting HC3 standard errors alongside conventional ones [26]. Standard weighted least squares results are used throughout [27].

3.3. Comparison models

Twelve models are compared: (M1) naïve; (M2) naïve with drift; (M3) Holt with additive trend [28]; (M4) ARIMA(0,1,1) [29]; (M5) linear regression on time; (M6) OLS with phase dummies; (M7) PA-WLS; (M8) standardised Ridge with phases and weights [30]; (M9) SVR with an RBF kernel; (M10) Random Forest; (M11) Gradient Boosting; and (M12) XGBoost. The machine-learning models use the same feature matrix [t, D_C, D_R]. SVR is trained on z-score standardized features within a standardization–regression pipeline that is re-estimated at every validation origin; the tree-based models are invariant to rescaling and keep the features on their original scale.

Ridge is retained as a comparison model but not as the main specification, for two reasons. The variance inflation factors of t, D_C and D_R are all below 2, so there is no evidence of severe multicollinearity; and the Ridge penalty depends on the scale of the variables, so that with the year variable on its original scale the same penalty shrinks the phase coefficients far more strongly than the trend coefficient. Figure 5(b) quantifies this: at α = 1 on unstandardized features, the 2026 forecast is already displaced by several

hundred students, whereas the standardized version is essentially flat over the same range. Standardized Ridge with α = 0.3, selected by weighted leave-one-out cross-validation, is therefore retained as M8, while PA-WLS remains the main model because its statistical interpretation is clearer.

Two features of the benchmark group must be stated for the results to be read correctly. First, when the sum of squared errors is minimized on this series, the two smoothing parameters of Holt converge to values close to zero; M3 therefore degenerates into a deterministic linear trend and coincides with M5 in both in-sample fit and the 2025 forecast. Second, ARIMA(0,1,1) is estimated by maximum likelihood in state-space form, so its in-sample statistics depend on filter initialization. SVR uses C = 10,000 and epsilon = 50; Random Forest uses 500 trees; Gradient Boosting and XGBoost each use 300 trees, a learning rate of 0.05, and a maximum depth of 2. The random seed is fixed at 42. Computations were carried out in Python using NumPy [31], SciPy [32], statsmodels [33], and scikit-learn [34]; library versions are provided in the data availability statement, as results for the tree-based models may vary slightly across versions.

3.4. Evaluation metrics and validation protocol

Four metrics are used: RMSE, which penalizes large deviations; MAE, readily interpreted in units of students; MAPE, computed over multiple validation points; and R^2_w for in-sample fit. One distinction is observed strictly: an absolute percentage error for a single year is an APE, not a MAPE. Accordingly, the 2025 result is reported as an APE, and MAPE is reserved for the set of eight rolling origins. This distinction follows the treatment of error measures in Hyndman and Koehler [35] and in Hyndman and Athanasopoulos [36].

The main protocol is rolling-origin one-step-ahead forecasting over the target years 2018–2025, giving eight validation origins. At each origin, the model is re-estimated on all data prior to the target year and forecasts exactly one year ahead, in line with Tashman [22]. To isolate the value of phase information, the study reports two scenarios. The oracle scenario assumes that the phase label for the target year is known at the time of forecasting; five of the eight origins then carry non-zero labels, while 2018, 2019, and 2025 have $D_C = D_R = 0$. This measures the maximum value phase information can provide. The ex ante scenario sets $D_C = D_R = 0$ for the coming year, reflecting an institution with no advance knowledge of a new shock. The distinction matters: a phase indicator identified after the event is not information available in real time.

4. RESULTS

4.1. Descriptive statistics and structural breaks

Over the whole series, the mean is 5,101 students, with a standard deviation of 1,530.7 and a coefficient of variation of approximately 30.0%. The baseline phase averages 4,449.0 students, the pandemic phase 7,414.7 students, and the restructuring phase 3,971.5 students. The year 2025 recorded 6,287 students, above the baseline mean but below the 2021 peak of 8,547. Figure 1 shows the full trajectory and the year-on-year changes, with the four regimes shaded so that the 2021 peak, the 2023 trough and the subsequent recovery are visually explicit; Table 2 gives the corresponding phase-level statistics.

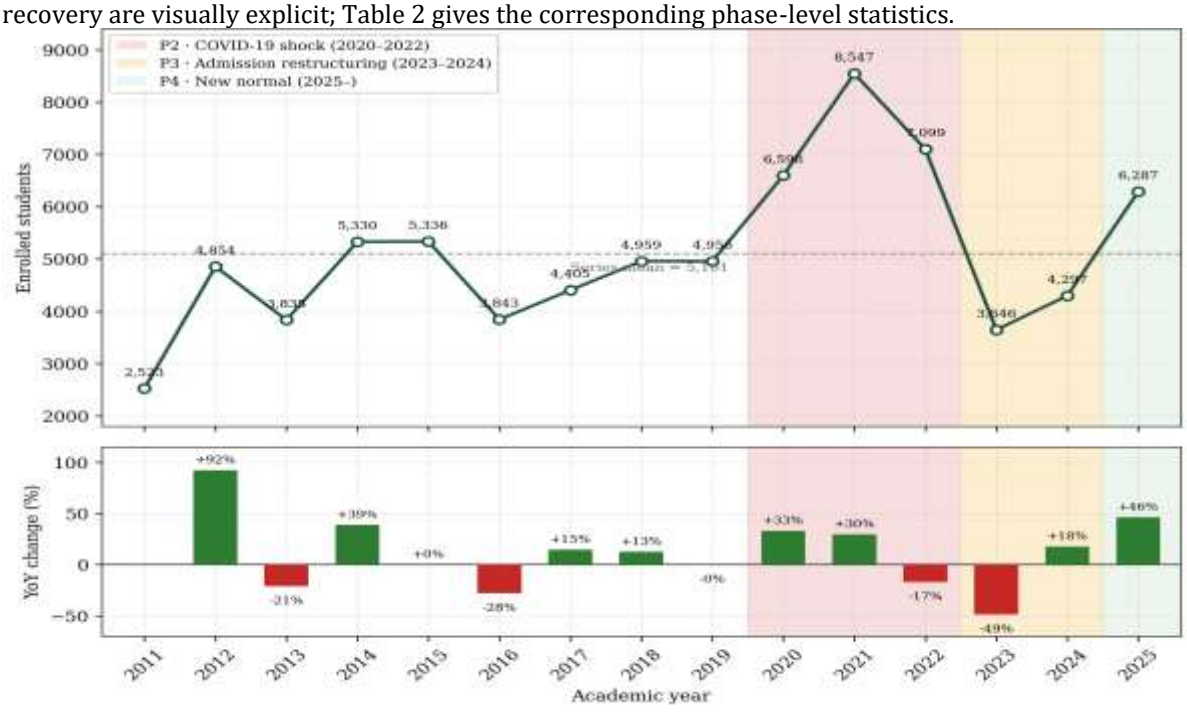

Figure 1. Annual first-year enrolment and year-on-year growth, 2011–2025.
Note. Shaded bands mark the pandemic (P2), restructuring (P3) and new-normal (P4) regimes; the dashed line is the series mean of 5,101 students. The lower panel gives year-on-year percentage change.

Table 2. Descriptive statistics by historical phase

Phase	n	Mean	SD	Min	Max	Relative to P1
P1 – Baseline (2011–2019)	9	4,449.0	915.1	2,523	5,336	—
P2 – Pandemic (2020–2022)	3	7,414.7	1,012.1	6,598	8,547	+66.7%
P3 – Restructuring (2023–2024)	2	3,971.5	460.3	3,646	4,297	−10.7%
P4 – New normal (2025)	1	6,287.0	—	6,287	6,287	+41.3%
Full series	15	5,101.0	1,530.7	2,523	8,547	—

Note. The final column expresses each phase mean as a percentage deviation from the P1 mean. P4 contains a single observation, so no standard deviation is defined.

Table 3 reports the trend and structural break tests. The battery comprises a one-way ANOVA together with Welch's unequal-variance t-test [37] for differences between phase means; the Chow test [19] for coefficient stability at two candidate break points; an F-test of the joint phase effect within PA-WLS; the Mann–Kendall test [38] for monotonic trend; and the augmented Dickey–Fuller [39] and KPSS [40] tests for a unit root and for stationarity around a constant. Three results deserve emphasis. The differences in means between P1, P2 and P3 are jointly significant. The Chow test does not confirm a break at 2020 at the 5% level but does confirm one at 2023. Neither the Mann–Kendall test nor the simple linear regression indicates a strong monotonic trend across the entire series, which supports a specification of a baseline trend plus phase shifts rather than a single trend line.

Table 3. Trend and structural break tests

Test	H_0	Statistic	p-value	Decision
ANOVA P1–P3	The phases have equal means	F = 13.710	0.001	Reject H_0
Welch t: P1 vs P2	The two phases have equal means	t = −4.499	0.018	Reject H_0
Welch t: P1 vs P3	The two phases have equal means	t = 1.070	0.358	Do not reject
Chow at 2020	Coefficients invariant across 2020	F = 2.711	0.110	Do not reject
Chow at 2023	Coefficients invariant across 2023	F = 7.098	0.011	Reject H_0
F-test PA-WLS: $\beta_2 = \beta_3 = 0$	No phase effects	F = 5.328	0.024	Reject H_0
Mann–Kendall	No monotonic trend	Z = 1.584	0.113	Do not reject
ADF	A unit root is present	−2.360	0.153	Do not reject
KPSS	Stationary around a constant	0.305	> 0.100	Do not reject

Note. The ADF test uses a constant-only specification with lag order 2; the KPSS test examines stationarity around a constant with a Newey–West window of 2. At n = 15, the power of both tests is very low, so a "do not reject" outcome must not be read as evidence in favor of the null. At the 2023 breakpoint, the second regression segment contains only three observations (2023–2025) for two parameters, so the Chow test there rests on very limited degrees of freedom and should be interpreted with caution.

4.2. Estimates and residual diagnostics

Table 4 reports the PA-WLS estimates on the full series. The model attains R^2_w = 0.666 (adjusted 0.575), F(3,11) = 7.314 with p = 0.006, AIC = 249.47, BIC = 252.31, RMSE = 692.7, MAE = 572.6 and an in-sample MAPE of 12.56%. AIC and BIC are computed from the log-likelihood of the weighted model, which includes the term ½·Σ ln ω_i, and count k = 4 regression parameters; other conventions yield different values.

Table 4. PA-WLS estimates over the full 2011–2025 series

Parameter	Interpretation	Coefficient	SE	t	p	95% CI
β_0	Baseline level in 2011	3,754.195	373.212	10.059	< 0.001	[2,932.8; 4,575.6]
β_1	Trend per year	175.721	58.955	2.981	0.013	[46.0; 305.5]
β_2	Pandemic level shift	1,903.261	849.389	2.241	0.047	[33.8; 3,772.8]
β_3	Restructuring level shift	−1,979.208	1,058.850	−1.869	0.088	[−4,309.7; 351.3]

Note. Standard errors and confidence intervals are conventional; HC3 alternatives are reported in the text where they differ materially. β_2 and β_3 are descriptive level shifts by phase, not causal effects.

The trend coefficient is significant at the 5% level, indicating that once the phase shifts are controlled for, baseline enrolment grows by about 176 students per year. The positive β_2 and negative β_3 reflect two level shifts of opposite sign. These are descriptive effects by phase, not causal relationships: β_3 has p = 0.088 under conventional standard errors, and phase P3 contains only two observations, so it cannot be read as implying that the restructuring caused a decline of exactly 1,979 students.

Figure 2 presents the residual and influence diagnostics, computed from the weight-standardized residuals, consistent with the WLS specification. The Durbin–Watson statistic [41] is 2.22, showing no troubling autocorrelation (2.40 on the raw residuals). The Shapiro–Wilk test [42] yields p = 0.465, so normality is not rejected. Cook's distance [43] for 2011 is 0.357, above the warning threshold 4/n = 0.267, and 2025 carries the highest leverage in the series (h = 0.64) because it is the only observation in phase P4 and lies at the right-hand boundary. Taken together, the diagnostics show a model stable enough to serve as a descriptive and controlled forecasting tool, but not one that should be presented as a highly accurate probabilistic model; the prediction intervals below remain intervals conditional on the specification.

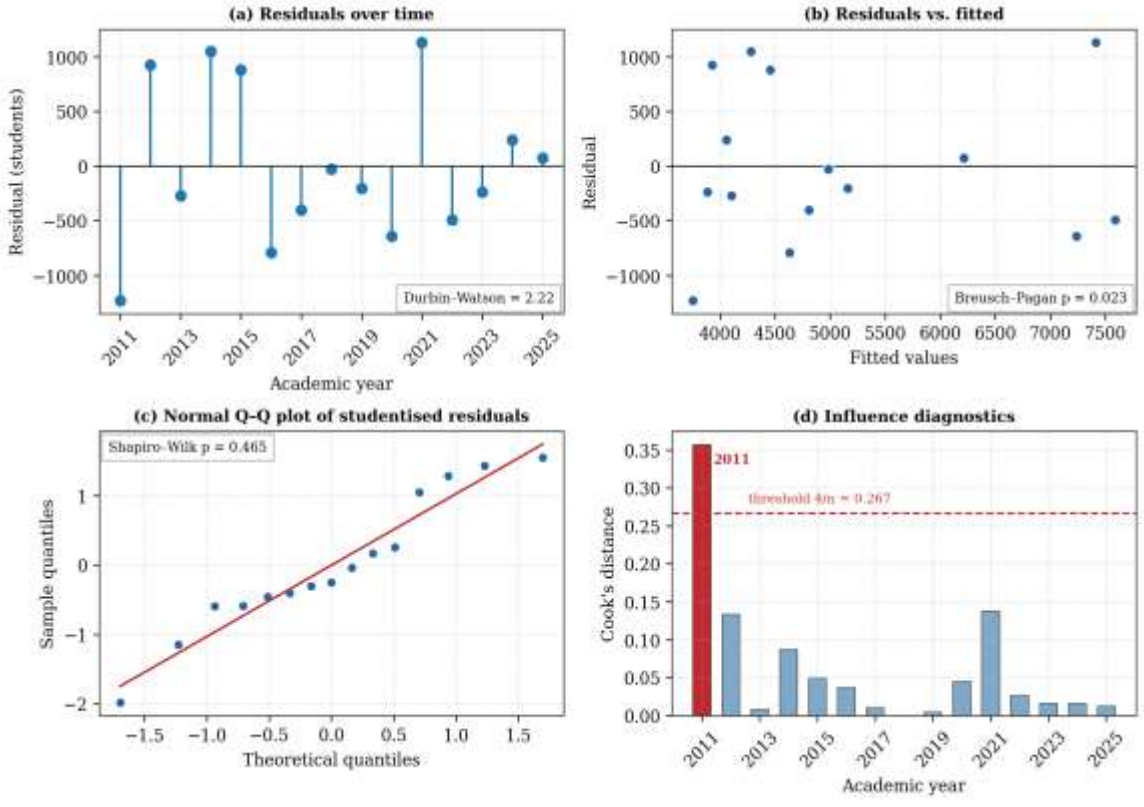

Figure 2. Residual and influence diagnostics for the phase-aware weighted model.
Note. Panels report residuals over time, residuals against fitted values, a normal Q–Q plot of studentized residuals, and Cook's distance with the 4/n reference line. All quantities are computed on weight-standardized residuals.

4.3. Model comparison and rolling-origin validation

Table 5 compares the in-sample and out-of-sample performance of the twelve candidates. The three regressions with phase dummies lead the group out-of-sample, with rolling-origin MAPE between 15.91% and 16.22% — a spread of only 0.31 percentage points, indicating that most of the benefit comes from the phase structure rather than from the choice among OLS, WLS, and Ridge. By contrast, Gradient Boosting attains an in-sample R^2 of 0.999 and an in-sample MAPE of 0.56%, yet its rolling-origin MAPE is 24.61%. Random Forest and XGBoost show the same pattern. This is a clear sign of overfitting at n = 15, compounded by the fact that tree-based models do not extrapolate beyond the training range and are therefore structurally unsuited to problems that require forecasts outside that range. Figure 3 makes the contrast clear: the tree ensembles track the observed series almost exactly, while the phase-aware model reproduces its level structure using only four parameters.

Table 5. In-sample and out-of-sample performance of the twelve models

Model	In-sample R^2	In-sample RMSE	In-sample MAPE (%)	RO RMSE	RO MAPE (%)	2025 forecast	APE 2025 (%)
M8 Standardised Ridge + phases + weights	0.779	696	12.57	1,150	15.91	5,979	4.90
M6 OLS + phase dummies	0.781	693	12.54	1,137	16.20	6,133	2.45
M7 PA-WLS	0.781	693	12.56	1,133	16.22	6,087	3.19
M9 SVR-RBF	0.759	726	10.52	1,393	19.81	5,233	16.76
M10 Random Forest	0.916	427	7.85	1,529	22.35	4,696	25.31
M11 Gradient Boosting	0.999	36	0.56	1,574	24.61	4,567	27.36
M12 XGBoost	0.831	608	8.36	1,795	26.88	5,958	5.23
M5 Linear regression on time	0.219	1,307	22.37	1,896	27.28	6,201	1.37
M1 Naïve	−0.111	1,559	25.01	1,775	27.60	4,297	31.65
M3 Holt	0.219	1,307	22.37	1,899	27.69	6,201	1.37
M2 Naïve + drift	−0.081	1,537	24.72	1,842	28.40	4,433	29.48
M4 ARIMA(0,1,1)	−0.182	1,608	30.72	1,901	31.65	4,949	21.29

Note. Models are ordered by rolling-origin (RO) MAPE. The in-sample R^2 and RMSE columns are computed without weights for all twelve models to ensure comparability; the value 0.781 for M7 therefore differs from R^2_w = 0.666 in Table 4, which is the weighted quantity defined in equation (5). RO columns are one-step-ahead results over eight origins under the oracle scenario. M3 coincides with M5 because the Holt smoothing parameters converge to values close to zero on this series (see Section 3.3).

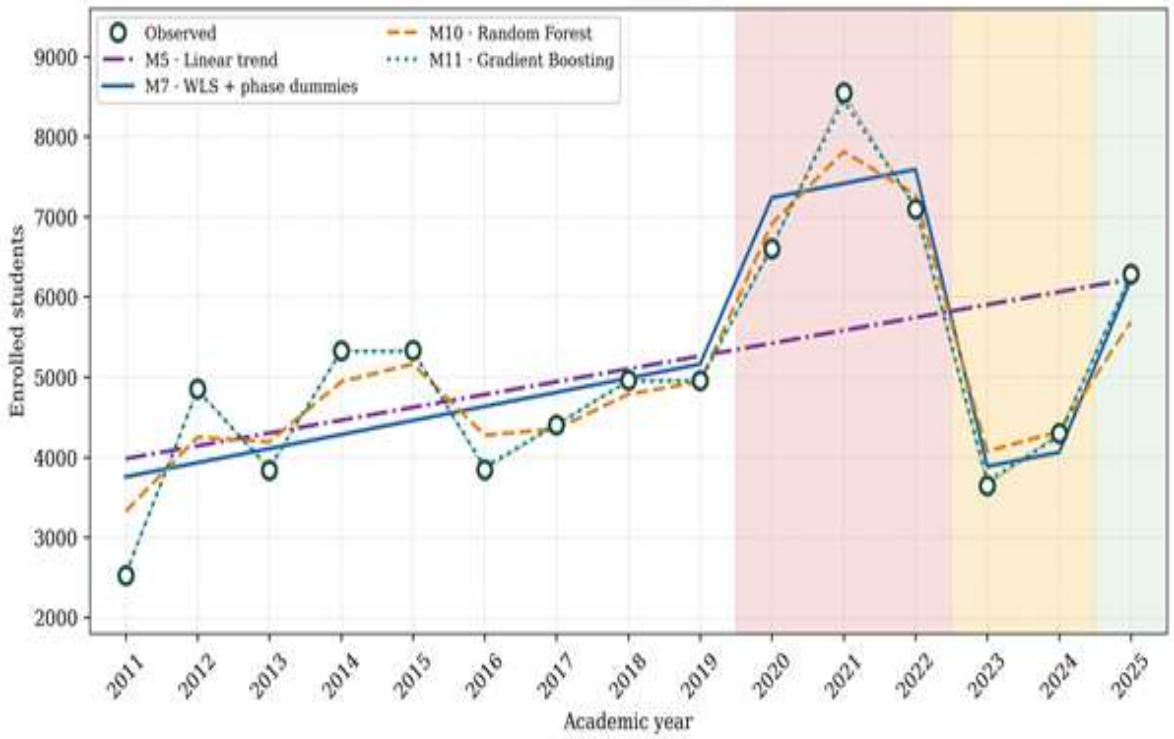

Figure 3. In-sample fitted values from the principal regression and tree-ensemble specifications.

Note. The near-exact fit of Gradient Boosting illustrates why in-sample accuracy alone is insufficient for a series of fifteen annual observations. Shaded bands mark the phases as in Figure 1.

To test whether the improvement of PA-WLS over the plain linear regression exceeds what chance would produce, three tests were applied to the eight pairs of errors from identical validation origins: the paired t-test on APEs gives $t = -1.437$ ($p = 0.194$); the Diebold–Mariano test on squared errors [44] gives $DM = -1.707$ ($p = 0.132$); and the Wilcoxon signed-rank test [45] gives $p = 0.438$. The 11.06-point difference in MAPE is therefore large in practical terms but does not reach significance at the 5% level with only eight origins.

Table 6 sets out the rolling-origin detail for PA-WLS. The largest error occurs in 2023 (APE 57.69%), precisely the year identified as the structural break: when forecasting 2023, the model had data only to 2022 and had never observed phase P3. Once data from the new phase became available, the APE fell to 11.37% in 2024 and 3.19% in 2025. The principal limitation of the model is therefore not the WLS estimator but its inability to recognize an entirely new phase. The final row also serves as the independent 2025 anchor: trained on 2011–2024, the model gives $\beta_0 = 3{,}793.991$, $\beta_1 = 163.752$, $\beta_2 = 1{,}983.153$, and $\beta_3 = -1{,}869.394$, and, with $D_C = D_R = 0$ and $t = 14$, forecasts 6,086.5 against an actual 6,287. That 3.19% is a favorable result but rests on a single point; it is not converted into a MAPE and does not constitute evidence that errors will be below 5% in later years. The rolling-origin MAPE of 16.22% remains the appropriate measure of average performance.

Table 6. One-step-ahead rolling-origin validation of PA-WLS

Target year	Training data	Actual	Forecast	Error	APE (%)
2018	2011–2017	4,959	5,035.9	−76.9	1.55
2019	2011–2018	4,956	5,180.5	−224.5	4.53
2020	2011–2019	6,598	5,257.3	+1,340.7	20.32
2021	2011–2020	8,547	6,759.7	+1,787.3	20.91
2022	2011–2021	7,099	7,821.7	−722.7	10.18
2023	2011–2022	3,646	5,749.4	−2,103.4	57.69
2024	2011–2023	4,297	3,808.5	+488.5	11.37

Target year	Training data	Actual	Forecast	Error	APE (%)
2025	2011–2024	6,287	6,086.5	+200.5	3.19

Note. Oracle scenario. At each origin, the model is re-estimated on all prior data. The 2025 row is the independent anchor-point validation discussed in the text and is not used to compute a MAPE on its own.

4.4. The value of phase information

Table 7 separates the two phase-information scenarios. When phase information is assumed available, the MAPE of PA-WLS is 16.22%; without it, the figure rises to 22.80%. The 6.58-point difference indicates that early recognition of the admission regime matters considerably. This is, however, the value of information within one particular series, and should not be read as implying that any early-warning system will reduce error by exactly 6.58 points.

Table 7. Performance under the two phase-information scenarios

Model	Scenario	RMSE	MAE	MAPE (%)
Plain linear regression	Phase-independent	1,895.9	1,316.5	27.28
OLS + phase dummies	With phase information (oracle)	1,137.1	865.9	16.20
OLS + phase dummies	Without phase information (ex ante)	1,613.3	1,261.3	22.69
PA-WLS	With phase information (oracle)	1,133.2	868.0	16.22
PA-WLS	Without phase information (ex ante)	1,617.2	1,272.0	22.80

Note. The oracle scenario assumes the phase label of the target year is known at the time of forecasting; the ex ante scenario sets both indicators to zero for the coming year. Both are computed over the same eight rolling origins.

4.5. Forecasts for 2026–2027

After the anchor-point validation, the model is re-estimated over the whole period 2011–2025. Both forecast years are assumed to be in the normal state, so $D_C = D_R = 0$. The prediction interval for a new observation is

$$PI_{95\%} = \hat{y}_0 \pm t_{0.975,\,n-p} \cdot s \cdot \sqrt{\frac{1}{\omega_0} + x_0'(X'\Omega X)^{-1}x_0}, \qquad \omega_0 = 1 \tag{6}$$

Table 8 reports the resulting forecasts, and Figure 4 places them in the context of the fitted historical series and the 2025 anchor. The central forecast rises by 1.64% in 2026 and 2.75% in 2027, but the 95% interval spans roughly ±2,100 students around the central value. The figure of 6,390 should therefore not be presented as a firm target; it is the center of a wide forecast distribution, and its width follows from the intrinsic noise of the series and the very small sample.

Table 8. Forecast enrolment for 2026–2027

Year	State	Point forecast	Forecast SE	80% PI	95% PI	Growth
2025	Actual – anchor point	6,287	—	—	—	—
2026	Stable normal	6,390	961.2	[5,079; 7,701]	[4,274; 8,506]	+1.64%
2027	Stable normal	6,566	998.5	[5,204; 7,927]	[4,368; 8,763]	+2.75%

Note. Intervals are computed from equation (6) and are conditional on the specification and on the assumption that 2026–2027 remain in the normal state. Section 5.4 discusses why that assumption may no longer hold.

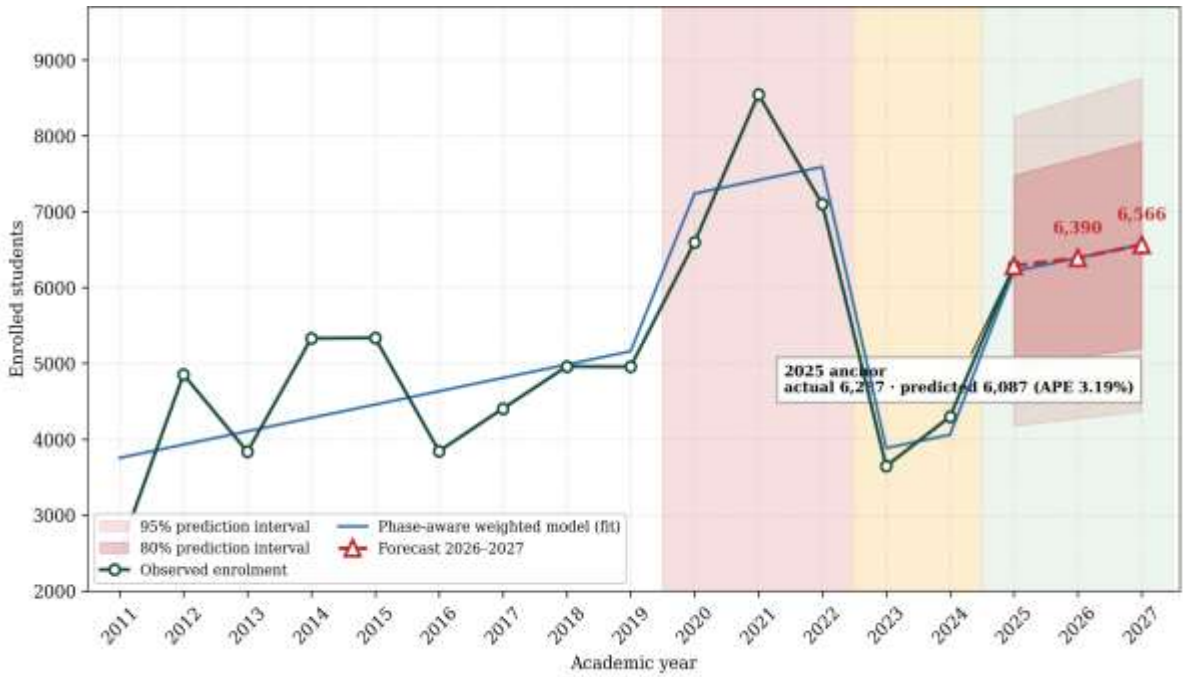

Figure 4. Phase-aware weighted forecasts for 2026–2027 with 80% and 95% prediction intervals.

Note. The 2025 anchor point is shown separately as an independent validation observation and is not included in the rolling-origin error calculation.

4.6. Sensitivity to weighting and to specification

Figure 5 evaluates the robustness of the 2026 forecast along three dimensions. Panel (a) varies the weight ω from 0.1 to 1.0 and moves the 2026 forecast only from 6,384.9 to 6,407.8, a total range of about 0.36%; the coefficients β_1, β_2, and β_3 likewise shift by less than one percent. Weighting is therefore not the decisive component of the forecast. Panel (b) shows the scale dependence of the Ridge penalty discussed in Section 3.3. Panel (c) is the more consequential result: across six defensible specifications, the 2026 forecast ranges from 6,001 students (excluding the influential 2011 observation) to 6,861 (a log-linear model with smearing retransformation), with OLS plus phases at 6,408, standardized Ridge at 6,327, and a Holt or linear trend at 6,382.

That range of 860 students is about 13.5% of the central forecast and roughly thirty-seven times the range induced by the weighting parameter. Specification uncertainty is thus of the same order as, and by this measure larger than, the statistical uncertainty conveyed by the standard error — a comparison that should accompany any single point forecast derived from a series of this length.

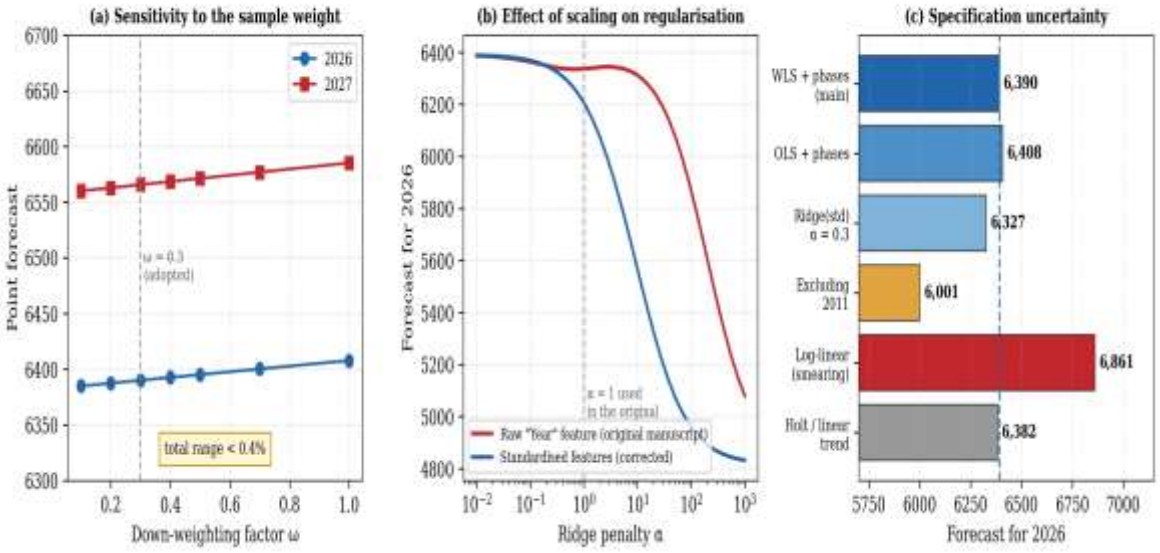

Figure 5. Sensitivity of the 2026 forecast to weighting, regularisation and model specification.
Note. Panel (a) varies the down-weighting factor ω; panel (b) contrasts the Ridge regularisation path on unstandardized and standardized features; panel (c) compares point forecasts across six alternative specifications, with the dashed line marking the main PA-WLS forecast of 6,390. The small effect of weighting in panel (a) is separated from the substantially larger effect of specification choice in panel (c).

5. DISCUSSION

5.1. Specification matters more than algorithm

The central finding is not the number 6,390. It is that a single linear trend line does not adequately describe the series, and that correcting the specification does more for forecast accuracy than any increase in algorithmic capacity. When the phases are ignored, the trend coefficient reaches only $p = 0.079$ with $R^2 = 0.219$; when the phase indicators are included, it becomes 175.72 with $p = 0.013$. The three phase-aware regressions differ by 0.31 percentage points in MAPE, while the gap to the plain linear trend is 11.06 percentage points, so the phase structure — not the estimator — drives the improvement. The same conclusion emerges from the weighting analysis: if the phase indicators already absorb most of the level shift, down-weighting becomes a secondary adjustment. Modeling the change matters more than reducing its influence.

The failure of the tree ensembles is a textbook illustration of the distinction between fit and forecast. With fifteen points, a model of high representational capacity can memorize the historical observations without learning any extrapolating regularity, and tree-based learners cannot, in principle, predict outside the range of their training values. This should not be read as a general verdict on machine learning for enrolment forecasting: Shao et al. [5], Soltys et al. [6] and Al-Naymat and Al-Betar [7] all demonstrate its usefulness on record-level admission data. The conclusion here is narrower and conditional — for a univariate series of fifteen annual totals from a single institution, additional algorithmic complexity buys nothing.

The validation design is what makes this visible. On the 2025 point alone, the plain linear regression has an APE of 1.37%, better than the 3.19% of PA-WLS; selecting a model on that basis would invert the ranking obtained over eight origins, where the same regression records 27.28% against 16.22%. This is precisely why Tashman [22] recommends multiple origins with re-estimation, and why the honest reading of the comparison must also acknowledge that eight origins are too few to establish significance.

5.2. Interpreting the phase coefficients

The positive pandemic coefficient of +1,903.26 indicates that enrolment in 2020–2022 stood above the model's baseline level. This runs counter to the United States evidence of pandemic-period decline [15], but the divergence is not a methodological contradiction: a global shock can produce different responses depending on national context, market structure and policy. At least three mechanisms merit examination in further research — that the opportunity cost of study falls when the labor market weakens, that disruption to overseas study retains some demand domestically, and that changes in admission methods and eligibility criteria during the period widened the pool of eligible candidates. All three remain hypotheses, since the present dataset contains no variables by which to identify them. The indicator D_C is not a COVID-19 variable in any causal sense; it labels a period in which many things changed at once, and the paper does not conclude that the pandemic increased enrolment by 1,903 students.

The restructuring coefficient of −1,979.21 similarly indicates that enrolment in 2023–2024 fell below the baseline level, and the Chow break at 2023 supports treating this as a regime change. Institutionally, the period followed the early-admission provision of Circular 08/2022 and preceded the repeal of Article 18 by Circular 06/2025 [16,17]. But phase P3 contains only two observations; β_3 carries $p = 0.088$ under conventional standard errors, and the smaller p-value obtained under HC3 does not remove the limitation imposed by the number of observations. The defensible conclusion is that the data provide evidence of a level shift around 2023, without precisely determining its magnitude or cause.

5.3. Managerial implications

Because the 95% interval for 2026 runs from 4,274 to 8,506 students, a resource plan anchored on 6,390 alone would ignore most of the risk. Forecasts of this kind are better converted into planning thresholds: the central value for the base plan, the lower bound of the 80% interval — about 5,079 students — to test program viability and the balance of resources, and the upper bound of about 7,701 to test classroom, dormitory and laboratory capacity together with the supply of visiting lecturers.

Two further implications follow from the results. The 6.58-point gap between the oracle and ex ante scenarios suggests that investment in input information may be worth more than a change in algorithm: earlier monitoring of application volumes, policy signals, and admission-market conditions would allow the model to recognize a new state sooner. And the 57.69% error at the 2023 break shows what such a

model cannot do. It cannot forecast a regime it has never observed, so a practical forecasting system needs a regime-change alerting layer sitting outside the trend model rather than a more elaborate trend model. Both propositions are managerial hypotheses to be tested with higher-frequency data.

5.4. Limitations

- The sample contains fifteen observations. The low residual degrees of freedom reduce the power of every test reported here and make model comparison susceptible to bias; the difference between PA-WLS and the linear trend does not reach significance at the 5% level.
- The series contains a single outcome variable. The separate effects of tuition fees, cut-off scores, regional competition, applicant pool size and application behavior cannot be identified.
- The phase indicators are defined from known history. A new, previously unseen shock would sharply increase error, as it did in 2023.
- Admission regulation has continued to change after the study period. Circular 06/2026/TT-BGDĐT of 15 February 2026 promulgated an entirely new regulation [18], so the assumption in Table 8 that 2026–2027 remain in a stable normal state may no longer be appropriate and should be revisited once the 2026 figures are available.
- The year 2011 is highly influential by Cook's distance, and excluding it moves the 2026 forecast to 6,001 students. The results are sensitive to a small number of observations at the start of the series.
- The Breusch–Pagan test indicates changing variance. HC3 standard errors support inference about the coefficients but do not resolve all uncertainty in the prediction intervals.
- The phase coefficients belong to one institution. Only the procedure, not the values of β_2 and β_3, transfers to other institutions.

6. CONCLUSIONS AND FURTHER WORK

The 2011–2025 enrolment series should not be treated as a homogeneous linear trend. There is statistical evidence of a break in 2023 and substantial differences between phases, and the PA-WLS model — a re-based time variable with two-level indicators and reduced weights on the volatile years — better describes the data structure than a plain linear regression. The rolling-origin MAPE falls from 27.28% to 16.22%, though with only eight origins this 11.06-point difference does not reach significance at the 5% level (paired t: $p = 0.194$; Diebold–Mariano: $p = 0.132$). The study therefore does not claim that PA-WLS is optimal or universally superior; the well-founded conclusion is that phase information has demonstrable empirical value in this series.

None of the nonlinear machine-learning algorithms outperforms the phase-aware regressions. The tree ensembles achieve very high in-sample fit and forecast worst out of sample, which illustrates the risk of overfitting in small samples as clearly as any constructed example could. Re-estimated on the whole series, the model gives central forecasts of 6,390 students for 2026 and 6,566 for 2027, with 95% prediction intervals of [4,274; 8,506] and [4,368; 8,763], and the specification analysis places the 2026 figure within the range 6,001–6,861, depending on modeling choices. The results are consequently suited to scenario-based planning rather than to fixing a single enrolment target.

Four directions for further work follow, in order of expected return. Extend the data dimensions before increasing model complexity: applications submitted, cut-off scores, relative tuition fees, the number of upper-secondary graduates in the catchment area, and measures of competitive intensity. Increase the frequency of observation from annual totals to admission-round or weekly data during the admission season, which is the precondition for evaluating more complex nonlinear models at all. Disaggregate by program or program group, build a hierarchical model, and reconcile to the institution-wide total. And develop a monitoring system that updates forecasts as new data arrive and raises an alert when observations fall outside the forecast range — the layer that would have caught 2023. Deep learning models should be considered only after the number of observations and the data dimensionality have been substantially increased; at $n = 15$, fitting a network with thousands of parameters is not justified.

DECLARATIONS AND ACKNOWLEDGEMENTS

Data availability. The aggregate series for 2011–2025 is reported in full in Table 1. The data are institution-level aggregates from Thu Dau Mot University, contain no student-identifying information, and can be made available on reasonable request. Analysis code and the library versions used are available from the corresponding authors.

Research ethics. The data contain no directly or indirectly identifying personal information.

Conflict of interest. The authors declare no conflict of interest.

Use of artificial intelligence tools. Computational and editorial support tools may have been used in data processing and language checking. The research design, choice of methods, verification of results, interpretation, and scientific responsibility rest with the authors.

Funding. This article is a product of an institutional-level research project (Code: DT.22.3-062).

REFERENCES

[1] Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, 46(5), 31–40. https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

[2] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery, 10(3), e1355. https://doi.org/10.1002/widm.1355

[3] Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. Telematics and Informatics, 37, 13–49. https://doi.org/10.1016/j.tele.2019.01.007

[4] Parajuli, D., Vo, D. K., Salmi, J., & Tran, N. T. A. (2020). Improving the performance of higher education in Vietnam: Strategic priorities and policy options. Washington, DC: World Bank. https://doi.org/10.1596/33681

[5] Shao, L., Ieong, M., Levine, R. A., Stronach, J., & Fan, J. (2022). Machine learning methods for course enrollment prediction. Strategic Enrollment Management Quarterly, 10(2), 11–29. ERIC: EJ1356234. https://eric.ed.gov/?id=EJ1356234

[6] Soltys, M., Dang, H., Reyes Reilly, G., & Soltys, K. (2021). Enrollment predictions with machine learning. Strategic Enrollment Management Quarterly, 9(2), 11–18. ERIC: EJ1311204. https://eric.ed.gov/?id=EJ1311204

[7] Al-Naymat, G., & Al-Betar, M. A. (2024). University student enrollment prediction: A machine learning framework. In K. Daimi & A. Al Sadoon (Eds.), Proceedings of the Third International Conference on Innovations in Computing Research (ICR'24). Lecture Notes in Networks and Systems, vol. 1058 (pp. 51–62). Cham: Springer. https://doi.org/10.1007/978-3-031-65522-7_5

[8] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. PLOS ONE, 13(3), e0194889. https://doi.org/10.1371/journal.pone.0194889

[9] 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

[10] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting, 38(4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

[11] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

[12] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[13] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[14] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York: ACM. https://doi.org/10.1145/2939672.2939785

[15] National Science Board, National Science Foundation. (2022). Higher education in science and engineering. Science and Engineering Indicators 2022 (NSB-2022-3). Alexandria, VA. https://ncses.nsf.gov/pubs/nsb20223/

[16] Ministry of Education and Training [Vietnam]. (2022). Circular No. 08/2022/TT-BGDĐT of 6 June 2022 promulgating the Regulation on university admission and on college admission for preschool education. Hanoi: MOET.

[17] Ministry of Education and Training [Vietnam]. (2025). Circular No. 06/2025/TT-BGDĐT of 19 March 2025 amending and supplementing a number of articles of the Regulation on university admission and on college admission for preschool education issued together with Circular No. 08/2022/TT-BGDĐT. Hanoi: MOET. Effective from 5 May 2025.

[18] Ministry of Education and Training [Vietnam]. (2026). Circular No. 06/2026/TT-BGDĐT of 15 February 2026 promulgating the Regulation on admission to undergraduate training disciplines and to the college-level discipline of preschool education. Hanoi: MOET.

[19] Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica, 28(3), 591–605. https://doi.org/10.2307/1910133
[20] Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. International Journal of Epidemiology, 46(1), 348–355. https://doi.org/10.1093/ije/dyw098
[21] Wagner, A. K., Soumerai, S. B., Zhang, F., & Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. Journal of Clinical Pharmacy and Therapeutics, 27(4), 299–309. https://doi.org/10.1046/j.1365-2710.2002.00430.x
[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] Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. Information Sciences, 191, 192–213. https://doi.org/10.1016/j.ins.2011.12.028
[24] Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis, 120, 70–83. https://doi.org/10.1016/j.csda.2017.11.003
[25] Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica, 47(5), 1287–1294. https://doi.org/10.2307/1911963
[26] Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. The American Statistician, 54(3), 217–224. https://doi.org/10.1080/00031305.2000.10474549
[27] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis (6th ed.). Hoboken, NJ: Wiley.
[28] Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting, 20(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015
[29] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Hoboken, NJ: Wiley.
[30] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634
[31] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2
[32] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2
[33] Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In Proceedings of the 9th Python in Science Conference (pp. 92–96). https://doi.org/10.25080/Majora-92bf1922-011
[34] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://www.jmlr.org/papers/v12/pedregosa11a.html
[35] 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
[36] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). Melbourne: OTexts. https://OTexts.com/fpp3
[37] Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved. Biometrika, 34(1–2), 28–35. https://doi.org/10.1093/biomet/34.1-2.28
[38] Mann, H. B. (1945). Nonparametric tests against trend. Econometrica, 13(3), 245–259. https://doi.org/10.2307/1907187
[39] Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366), 427–431. https://doi.org/10.2307/2286348
[40] Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, 54(1–3), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y
[41] Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. Biometrika, 38(1–2), 159–177. https://doi.org/10.1093/biomet/38.1-2.159
[42] Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

[43] Cook, R. D. (1977). Detection of influential observations in linear regression. Technometrics, 19(1), 15–18. https://doi.org/10.1080/00401706.1977.10489493
[44] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599
[45] Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin, 1(6), 80–83. https://doi.org/10.2307/3001968