EDUMATIC Jurnal Pendidikan Informatika

e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Volume 10, Issue 1, **70-79.**
https://e-journal.hamzanwadi.ac.id/index.php/edumatic
DOI: 10.29408/edumatic.v10i1.33925

Beyond Predictive Accuracy: Enhancing Parameter Stability in Multicollinear Time Series Forecasting via Regularisation

Daffa Kumara Khiar Faisa [1], Abu Salam [1,*]

[1] Universitas Dian Nuswantoro, Indonesia
[*] Corresponding author: Abu Salam, Universitas Dian Nuswantoro, Indonesia.
abu.salam@dsn.dinus.ac.id

Received: 15 January 2025 | Revised: 27 January 2026 | Accepted: 23 February 2026 | Published: 4 March 2026

Abstract

Multicollinearity in feature-based time series regression arises as a structural consequence of lagged and rolling feature construction. However, existing studies on Ridge and ElasticNet regularization adopt an accuracy-driven evaluation paradigm, with limited attention to parameter stability, shrinkage behavior, and sensitivity to regularization strength. This study shifts the evaluation of regularized linear models from predictive accuracy toward stability-oriented assessment. Using daily electricity consumption data from the UCI Repository, Linear Regression, Ridge, and ElasticNet models are examined under engineered temporal features derived from stability-based lag pruning, rolling statistics, and correlation-informed feature selection. Model evaluation focuses on bias–variance behavior, coefficient shrinkage, regularization sensitivity, and training–testing performance gaps. The results show that regularization improves stability, with the performance gap decreasing from 0.0961 in Linear Regression to 0.0608 under ElasticNet. These comparisons show that regularization stabilizes regression models via distinct shrinkage mechanisms, informing model selection beyond accuracy. Ridge exhibits conservative shrinkage averaging 6.06%, whereas ElasticNet induces stronger shrinkage averaging 46.32% and shows higher sensitivity to penalty strength. These findings provide methodological evidence that regularization in feature-based time series regression should be treated as a stability strategy rather than an accuracy optimization tool, offering guidance for electricity load forecasting under structurally redundant temporal features.

Keywords: elasticnet regularization; electricity consumption; multicollinearity; ridge regression; time series regression

To cite this article: Faisa, D. K. K., & Salam, A. (2026). Beyond Predictive Accuracy: Enhancing Parameter Stability in Multicollinear Time Series Forecasting via Regularisation. *Edumatic: Jurnal Pendidikan Informatika*, *10*(1), 70–79. https://doi.org/10.29408/edumatic.v10i1.33925

INTRODUCTION

Linear regression remains widely used in energy forecasting due to its transparent parameter structure and interpretability, which are critical for operational planning and policy analysis. However, when temporal feature engineering introduces structurally induced multicollinearity, parameter estimates become statistically unstable despite often maintaining competitive predictive accuracy. This creates a methodological paradox in time series forecasting: models may appear reliable based on accuracy metrics while producing fragile and inconsistent coefficient estimates, thereby undermining interpretability and potentially misleading policy decisions. However, representing temporal dependence through lagged variables and rolling statistical features inevitably introduces strong correlations among predictors, exposing linear regression models to substantial multicollinearity that threatens parameter stability and model reliability.

In time series models with correlated lagged predictors, multicollinearity inflates estimator variance and increases the sensitivity of coefficient estimates to small perturbations in the data, thereby undermining the reliability of linear regression models (Akhtar & Alharthi, 2024; Al-Essa et al., 2024; Alharthi & Akhtar, 2025), a concern that becomes operationally significant in daily electricity consumption forecasting with temporally engineered features. In feature-based time series regression, multicollinearity is not incidental but structural. Incidental multicollinearity typically arises from accidental correlations among regressors in cross-sectional settings and may be mitigated through variable selection or model respecification. In contrast, structural multicollinearity emerges inherently from overlapping lag windows, rolling statistics, and repeated temporal aggregation, where predictors are mechanically constructed from shared historical information (Basagaña & Barrera-Gómez, 2022; Demateis et al., 2024). Consequently, coefficient instability should be viewed as an inherent property of temporally engineered regression models rather than a consequence of poor model specification.

In daily electricity consumption forecasting, the use of lagged variables and rolling statistical features systematically induces structural multicollinearity, which directly manifests as coefficient instability in linear regression models. In the context of electricity consumption forecasting, coefficient instability is not merely a theoretical concern. Unstable parameter estimates can propagate into inaccurate peak load estimation, unreliable demand sensitivity analysis, and misleading operational decisions, potentially leading to financial inefficiencies and increased risk in grid management. These considerations indicate that, for energy applications, model reliability depends not only on predictive accuracy but also on the stability and robustness of estimated parameters.

Regularization techniques have been proposed to mitigate multicollinearity by constraining coefficient magnitudes. Ridge regression applies an L2 penalty that reduces variance through continuous shrinkage, while ElasticNet combines L1 and L2 penalties to balance shrinkage and sparsity under strong predictor correlation (Goulet Coulombe et al., 2022; Tay et al., 2023). Theoretically, regularization can be interpreted as a form of constrained optimization that deliberately introduces bias in exchange for variance reduction and more stable, practically reasonable parameter estimates, rather than merely as an accuracy enhancement mechanism (Guan & Burton, 2022; Sztepanacz & Houle, 2024). From a statistical perspective, shrinkage-based estimators reduce coefficient variance and quadratic risk under multicollinearity, thereby improving estimator stability rather than merely predictive accuracy (Al-Momani et al., 2025).

Despite the established theoretical link between multicollinearity and coefficient instability, empirical research in electricity consumption forecasting continues to prioritise predictive accuracy as the dominant evaluation objective. Regularized linear regression models are typically assessed using accuracy-centric metrics such as RMSE or MAE, and are frequently deemed successful based primarily on test performance (Koukaras et al., 2024; Uniejewski, 2024). Parameter stability, when acknowledged, is generally treated as a secondary diagnostic rather than as a central evaluation criterion. To date, stability has rarely been operationalised as a primary objective in feature-based time series regression. This accuracy-driven evaluation paradigm persists despite methodological evidence that estimator stability constitutes an independent scientific objective under correlated predictors (Bodinier et al., 2023; Haftorn et al., 2023), and that inappropriate evaluation design can lead to spurious model selection even when apparent predictive accuracy is high (Wang & Ruf, 2022).

This study contributes conceptually by reframing regularisation in temporally engineered regression as a stability-control mechanism under structural multicollinearity, rather than solely as a bias–variance trade-off adjustment for predictive performance. It contributes from an evaluation perspective by systematically incorporating parameter stability as a core assessment dimension alongside predictive accuracy in feature-based time series regression. Methodologically, it provides an empirical analysis of shrinkage behaviour, sensitivity to regularisation strength, prediction variance, and generalisation dynamics under correlated temporal predictors.

METHOD

This study employs a quantitative experimental design to examine the stability of linear regression models under structural multicollinearity induced by temporal feature engineering. The focus is not predictive optimization, but assessing how regularization affects parameter robustness, coefficient shrinkage, and generalization in time series regression. The analysis uses daily temporal resolution, and findings are not intended to generalize to coarser scales. The end-to-end workflow is summarized in Figure 1.

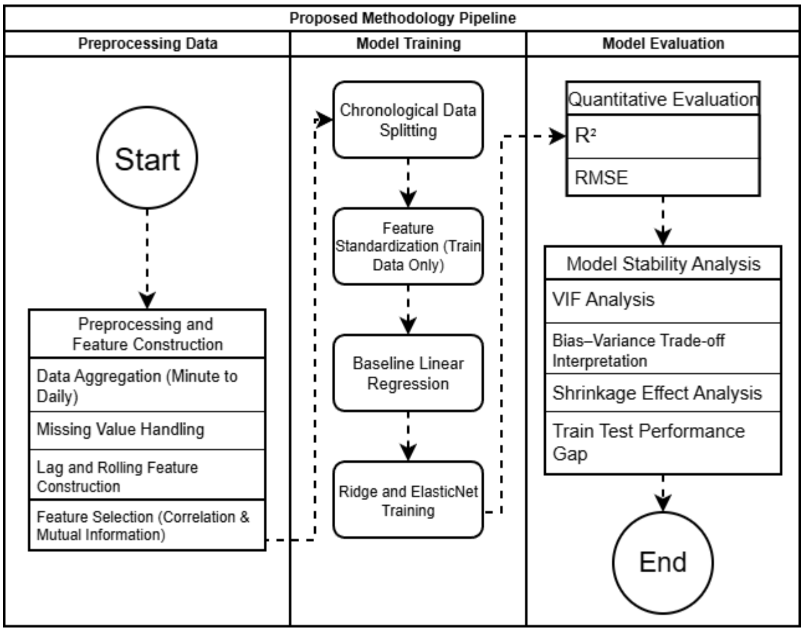

Figure 1. Proposed methodology pipeline illustrating preprocessing, model training, and evaluation stages

The population includes all electricity consumption records, while the analytical sample consists of daily aggregated observations. Daily aggregation preserves lag-induced temporal overlap relevant to stability analysis while excluding intra-day volatility. Chronological partitioning prevents information leakage, and robustness is evaluated using multiple train–test split ratios (70:30, 75:25, and 80:20), consistent with time series evaluation practices (Hewamalage et al., 2023). The impact of each preprocessing stage on dataset size is summarized in Table 1.

Research instruments include the electricity consumption dataset, a transparent preprocessing and feature engineering pipeline, and three regression models: Linear Regression, Ridge Regression, and ElasticNet. Feature scaling is performed using standardization parameters estimated solely from training data. Model validity is assessed using stability-oriented indicators rather than a single performance metric, including coefficient shrinkage magnitude, training–testing performance gaps, and prediction variance, which capture aspects of estimator stability grounded in penalized regression frameworks (Bodinier et al., 2023).

Table 1. Dataset conditions across preprocessing stages

Dataset Condition	Number of Rows	Information
Raw Dataset	2075259	Minute-level data, missing values present
After Daily Aggregation	1442	Daily frequency, missing values removed
After creating the lagged feature	1411	Daily data with lag features, NaN removed

Preprocessing includes daily aggregation and removal of missing observations (<5%) without interpolation to avoid temporal distortion. Correlation and mutual information are used for variable screening, followed by construction of lagged and rolling features, producing a high-dimensional temporal feature space. Structural multicollinearity is induced through basic and extended lag configurations. Stability-based pruning is applied in selected setups, while control configurations retain all features to isolate regularization effects. Regularization strength is explored across a logarithmic range of α values to examine shrinkage behavior and stability sensitivity.

Model performance is evaluated using RMSE and R^2. Perturbations include data partitioning, lag depth variation, and regularization strength exploration. Shrinkage magnitude is analyzed alongside prediction variance and train–test gaps to distinguish stabilizing contraction from over-penalization. All analyses use fixed random seeds, with sensitivity checks confirming robustness.

Coefficient shrinkage is quantitatively measured using Equation (1) to assess parameter stability across regularization settings, where $\beta_j^{(LR)}$ denotes the estimated coefficient from baseline Linear Regression, and $\beta_j^{(Reg)}$ represents the corresponding coefficient from the regularized model. This metric captures the relative magnitude of coefficient contraction induced by regularization and serves as an indicator of parameter sensitivity under multicollinearity.

$$Shrinkage_j(\%) = \frac{\left|\beta_j^{(LR)}\right| - \left|\beta_j^{(Reg)}\right|}{\left|\beta_j^{(LR)}\right|} \times 100 \tag{1}$$

RESULTS AND DISCUSSION

Results

The final daily dataset comprised 1,442 observations free of missing values and suitable for regression analysis. The electricity consumption series exhibits pronounced day-to-day variability, indicating strong temporal dependence relevant for feature-based regression modeling. Autocorrelation (ACF) and partial autocorrelation (PACF) analyses reveal statistically significant temporal dependence persisting across a wide range of lags, extending up to approximately 180 days. This persistence provides structural evidence that correlations among temporally engineered predictors are unavoidable rather than incidental. Based on stability frequency across validation folds, lag structures exhibit heterogeneous behavior, with several lag regions showing inconsistent selection. Following stability-based pruning and band grouping, three representative lag values, namely lag-1, lag-8, and lag-22, were retained. These lags capture short-term dependence, near-weekly recurrence, and weaker medium-range temporal effects, respectively, and serve as a stable reduced representation of the underlying temporal structure. Across validation folds, these three lags consistently appear with selection frequencies exceeding 0.8 across validation folds, indicating high selection stability, while higher-order lags exhibit substantially lower and unstable selection frequencies.

Correlation and mutual information analyses consistently identify Sub_metering_1, Sub_metering_2, and Sub_metering_3 as the most informative predictors. Global_intensity exhibits an almost perfect correlation with the target variable ($r = 0.999$) and is therefore excluded to avoid redundancy, while Voltage and Global_reactive_power display negligible correlation and mutual information contribution and are removed, as summarized in Table 2. The final feature set consists of Sub_metering_1, Sub_metering_2, and Sub_metering_3 combined with lagged terms (1, 8, and 22) and rolling statistical features (gap_roll7_mean, gap_roll7_std, and gap_roll14_mean). Variance Inflation Factor (VIF) analysis conducted on the baseline Linear Regression model indicates that multicollinearity remains present, particularly among rolling features due to overlapping temporal windows, with VIF values for rolling mean features exceeding conventional thresholds. In the baseline Linear Regression model, residual multicollinearity inflates coefficient variability by driving the design matrix toward near-singularity, a condition under which Linear Regression estimates become unstable as the optimisation prioritises minimisation of the residual sum of squares.

Table 2. Feature relevance evaluation based on correlation and mutual information (mi)

Variable	Correlation (r)	Mutual Infromation (MI)	Decision
Global_intensity	0.999	3.241	Removed
Sub_metering_1	0.721	0.559	Selected
Sub_metering_2	0.472	0.302	Selected
Sub_metering_1	0.532	0.258	Selected
Voltage	0.138	0.046	Removed
Global_reactive_power	0.053	0.086	Removed

Model performance results indicate distinct generalisation behavior across the three regression models. Linear Regression achieves an R^2 of 0.8542 on the training set and 0.7582 on the test set, yielding a train–test performance gap of 0.0961. Ridge Regression reduces this gap to 0.0805, while ElasticNet further narrows it to 0.0608, with numerical performance comparisons reported in Table 3. These reductions remain stable across alternative train–test split ratios (70:30, 75:25, and 80:20), with ElasticNet gaps remaining within a narrow range of 0.058–0.063 and Ridge gaps within 0.078–0.083, and no reversal in the relative stability ordering of the models is observed. Although ElasticNet achieves the highest test performance (R^2_test = 0.7852; RMSE_test = 0.1436), the primary contribution of these results lies not in the absolute magnitude of accuracy improvement, but in the systematic reduction of generalisation discrepancies under correlated predictors, indicating improved control of overfitting through regularization. Across all evaluated models, improvements in R^2 are consistently accompanied by corresponding reductions in RMSE, with no conflicting trends observed between accuracy metrics. This consistency across evaluation measures indicates that the observed performance gains under regularization are robust and not driven by metric-specific artefacts or isolated error patterns.

Table 3. Comparison of regression model performance on training and testing data

Model	R^2_train	R^2_test	RMSE_train	RMSE_test
Linear Regression	0.8542	0.7582	0.1616	0.1524
Ridge Regression	0.8520	0.7715	0.1628	0.1481
ElasticNet Regression	0.8460	0.7852	0.1661	0.1436

Regularization path analysis reveals contrasting stability behavior between Ridge Regression and ElasticNet. In Ridge Regression, increasing regularization strength (α) leads to a gradual decline in training performance, while test performance remains relatively stable across a broad penalty range (approximately $\alpha = 10^{-4}$ to 10^{-1}) before deteriorating under stronger penalization, as shown in Figure 3. For Ridge and ElasticNet, moderate penalties mark an inflection region of maximal generalization stability, beyond which increasing α induces underfitting via excessive bias. In contrast, ElasticNet exhibits substantially higher sensitivity to regularization strength, with test performance improving within a moderate penalty range (approximately $\alpha = 10^{-3}$ to 10^{-2}) before deteriorating sharply as α increases further, accompanied by rapid degradation in both training and testing performance, as illustrated in Figure 4. This non-linear response indicates a narrower stability region for ElasticNet relative to Ridge Regression.

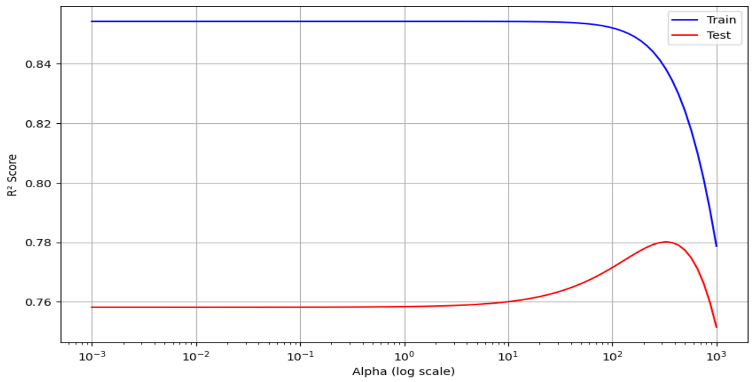

Figure 2. Trade-off curves between regularization strength (α) and R^2 scores for the ridge regression

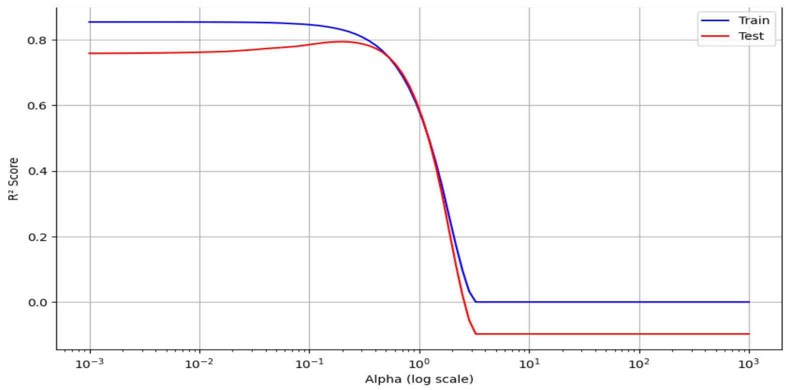

Figure 3. Trade-off curves between regularization strength (α) and R^2 scores for the elasticnet

Coefficient shrinkage analysis further differentiates the stability behavior of the two regularized models. Ridge Regression demonstrates moderate and relatively uniform coefficient contraction, with an average shrinkage of 6.06%, whereas ElasticNet induces substantially stronger shrinkage, with an average coefficient reduction of 46.32%, as illustrated in Figure 5. In the context of electricity consumption data with strong temporal autocorrelation, the substantially stronger shrinkage induced by ElasticNet reflects effective removal of redundant temporal information through its L1 component, rather than indiscriminate loss of informative signal. Consistent with comparisons between full temporal feature sets and stability-based pruning configurations, rolling statistical features exhibit stronger coefficient shrinkage than discrete lagged predictors due to overlapping temporal windows, while short-term lags such as lag-1 remain relatively stable and longer lags become increasingly sensitive under regularization.

This effect is particularly pronounced for ElasticNet, which exhibits substantially stronger shrinkage in the extended-lag configuration compared to the basic lag setting, reflecting its sensitivity to increased temporal overlap. In contrast, Ridge regression maintains relatively stable and gradual shrinkage patterns across both lag configurations, suggesting greater robustness to increasing lag-induced multicollinearity. Output stability, assessed through prediction variance, exhibits a consistent downward trend under regularization, decreasing from 0.1204 in Linear Regression to 0.1110 in Ridge Regression and further to 0.1023 in ElasticNet, with variations remaining within ±0.005 across alternative split ratios and lag configurations. The concurrent reduction in prediction variance and generalisation gap indicates that regularization systematically improves estimation, generalisation, and output stability under temporally correlated predictors.

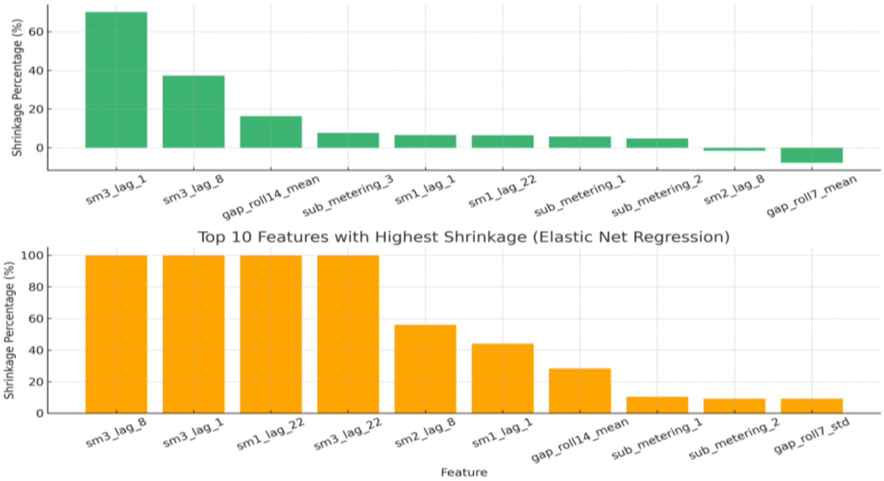

Figure 4. Coefficient shrinkage percentages of the ten most affected features under Ridge Regression and ElasticNet regularization.

Discussion

The findings demonstrate that regularisation improves estimation stability in linear regression models subjected to structural multicollinearity induced by temporally engineered features. In this context, let X denote the design matrix constructed from lagged and rolling predictors. Overlapping temporal windows generate dense linear dependence among columns of X, rendering the cross-product matrix $X'X$ ill-conditioned and causing several of its eigenvalues to approach zero. Such behaviour is consistent with distributed lag and time-varying exposure models, in which multicollinearity arises as a modelling consequence rather than a dataset-specific anomaly (Basagaña & Barrera-Gómez, 2022; Kariya et al., 2024; Shirato et al., 2024). Under ordinary least squares estimation, inversion of $X'X$ amplifies variance along poorly conditioned directions, producing unstable and highly sensitive coefficient estimates.

Ridge regression mitigates this instability by modifying the estimator through L2 penalisation. Replacing $X'X$ with $X'X + \alpha I$ increases the smallest eigenvalues of the system and improves numerical conditioning, thereby reducing variance amplification under correlated predictors (Goulet Coulombe et al., 2022; Guan & Burton, 2022). This spectral adjustment stabilises coefficients while retaining the complete temporal predictor structure. ElasticNet extends this mechanism by combining L2-based conditioning improvement with L1-induced sparsity. The L1 component selectively attenuates redundant lagged and rolling features, leading to stronger shrinkage in configurations characterised by greater temporal overlap, consistent with established analyses of penalised regression paths (Sztepanacz & Houle, 2024; Tay et al., 2023). The empirical shrinkage patterns observed here therefore reflect two distinct stabilisation mechanisms: continuous variance control in Ridge and sparsity-driven dimensional contraction in ElasticNet. The increased sensitivity of ElasticNet to regularisation strength further illustrates the trade-off between redundancy attenuation and bias dominance under stronger penalisation (Akhtar & Alharthi, 2024).

The contribution of this study does not reside in establishing that regularisation reduces estimator variance, a result well documented in statistical learning and econometric theory. Its novelty lies in repositioning parameter stability as the primary evaluation objective in feature-based time series regression under structural multicollinearity. Rather than evaluating regularised models exclusively through predictive accuracy, the present analysis operationalises stability through shrinkage magnitude, sensitivity to penalty strength, prediction variance, and generalisation gaps. These reframing shifts evaluative emphasis from performance optimisation toward estimator robustness and conditioning behaviour in temporally engineered regression settings, aligning with methodological discussions that emphasise estimator stability as an independent scientific objective (Bodinier et al., 2023; Haftorn et al., 2023).

The stabilising role of regularisation identified in this study is conditional on feature design complexity. In temporal regression settings with limited lag depth and minimal overlap among predictors, structural multicollinearity is substantially attenuated, reducing the necessity of aggressive shrinkage. Under such parsimonious configurations, estimator differences are expected to reflect predictive flexibility rather than variance control. The present findings therefore indicate that the importance of regularisation increases with lag proliferation and rolling-feature density. Stability gains observed here arise from feature-induced redundancy rather than representing a universal property of time series regression.

Although the empirical application focuses on electricity consumption forecasting, the structural mechanisms described extend beyond the energy domain. Econometric models incorporating distributed lag structures, high-dimensional macroeconomic forecasting frameworks, and temporally structured exposure models frequently generate dense predictor overlap that induces design-matrix ill-conditioning (Ariens et al., 2022; Dertli et al., 2024). In such contexts, multicollinearity is embedded in model construction rather than arising from sampling irregularities. A stability-oriented evaluation framework may therefore be relevant wherever temporally engineered predictors introduce structural dependence across regressors.

The implications of these findings are twofold: they pertain to model governance and policy-sensitive forecasting environments. In energy systems where regression coefficients inform elasticity estimation, demand responsiveness analysis, and capacity planning decisions, unstable parameters can lead to misinterpretation and misguided operational strategies. Inadequate consideration of demand sensitivity can result in the inefficient allocation of reserves, the suboptimal scheduling of generation, or the inappropriate adjustment of tariffs. Incorporating stability indicators into evaluation protocols enhances forecast accountability by ensuring that policy decisions are not based on statistically unreliable parameter estimates. Ridge regression provides more conservative and stable coefficient behavior under persistent multicollinearity. On the other hand, ElasticNet offers stronger redundancy control when calibrated properly, though it is more sensitive to penalty selection.

Several limitations constrain the interpretation of these findings. First, the analysis is conducted at daily temporal resolution, and stability dynamics may differ under alternative aggregation scales. Second, the empirical evaluation relies on a single dataset, limiting external validity across distinct demand regimes or regulatory contexts (Hewamalage et al., 2023; Wang & Ruf, 2022). Third, the exclusive focus on linear regression restricts inference regarding nonlinear or ensemble-based approaches, which may exhibit distinct stability characteristics under structural multicollinearity. Accordingly, the stability-oriented framework proposed here should be interpreted as model- and context-specific rather than universally generalisable.

CONCLUSION

This study establishes regularisation as a structural stabilisation mechanism in feature-rich linear time series regression, where multicollinearity arises inherently from temporally engineered lagged and rolling predictors. The principal contribution lies in repositioning parameter stability as a primary evaluation objective, rather than treating it as a secondary diagnostic subordinate to predictive accuracy. By demonstrating how conditioning of the design matrix governs estimator robustness under structural redundancy, the findings clarify the distinct stabilisation roles of Ridge and ElasticNet within high-dimensional temporal settings. These conclusions apply specifically to linear models characterised by dense predictor overlap and do not automatically extend to parsimonious specifications or nonlinear architectures such as deep learning models. More broadly, the stability-oriented perspective advanced here is relevant to econometric, distributed lag, and high-dimensional regression contexts in which multicollinearity is embedded in model construction. Future research should examine stability behaviour under rolling-origin validation schemes, investigate Bayesian shrinkage formulations that quantify posterior parameter uncertainty, and analyse estimator sensitivity under regime shifts and structural breaks to further delineate the theoretical and empirical boundaries of stability-based evaluation.

REFERENCES

Akhtar, N., & Alharthi, M. F. (2024). A comparative study of the performance of new ridge estimators for multicollinearity: Insights from simulation and real data application. *AIP Advances*, *14*(11). https://doi.org/10.1063/5.0236631

Al-Essa, L. A., Ebrahim, E. A., & Mergiaw, Y. A. (2024). Bayesian regression modeling and inference of energy efficiency data: the effect of collinearity and sensitivity analysis. *Frontiers in Energy Research*, *12*. https://doi.org/10.3389/fenrg.2024.1416126

Alharthi, M. F., & Akhtar, N. (2025). Modified Two-Parameter Ridge Estimators for Enhanced Regression Performance in the Presence of Multicollinearity: Simulations and Medical Data Applications. *Axioms*, *14*(7), 527. https://doi.org/10.3390/axioms14070527

Al-Momani, M., Yüzbaşı, B., Bataineh, M. S., Abdallah, R., & Moideenkutty, A. (2025). Shrinkage Approaches for Ridge-Type Estimators Under Multicollinearity. *Mathematics*, *13*(22). https://doi.org/10.3390/math13223733

Ariens, S., Adolf, J. K., & Ceulemans, E. (2022). Collinearity Issues in Autoregressive Models with Time-Varying Serially Dependent Covariates. *Multivariate Behavioral Research*, *58*(4), 687–705. https://doi.org/10.1080/00273171.2022.2095247

Basagaña, X., & Barrera-Gómez, J. (2022). Reflection on modern methods: Visualizing the effects of collinearity in distributed lag models. *International Journal of Epidemiology*, *51*(1), 334–344. https://doi.org/10.1093/ije/dyab179

Bodinier, B., Filippi, S., Nøst, T. H., Chiquet, J., & Chadeau-Hyam, M. (2023). Automated calibration for stability selection in penalised regression and graphical models. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, *72*(5), 1375–1393. https://doi.org/10.1093/jrsssc/qlad058

Demateis, D., Keller, K. P., Rojas-Rueda, D., Kioumourtzoglou, M. A., & Wilson, A. (2024). Penalized distributed lag interaction model: Air pollution, birth weight, and neighborhood vulnerability. *Environmetrics*, *35*(4). https://doi.org/10.1002/env.2843

Dertli, H. I., Hayes, D. B., & Zorn, T. G. (2024). Effects of multicollinearity and data granularity on regression models of stream temperature. *Journal of Hydrology*, *639*, 131572. https://doi.org/10.1016/j.jhydrol.2024.131572

Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, *37*(5), 920–964. https://doi.org/10.1002/jae.2910

Guan, X., & Burton, H. (2022, December). Bias-variance tradeoff in machine learning: Theoretical formulation and implications to structural engineering applications. *Structures 46*, 17-30. Elsevier. https://doi.org/10.1016/j.istruc.2022.10.004

Haftorn, K. L., Romanowska, J., Lee, Y., Page, C. M., Magnus, P. M., Håberg, S. E., Bohlin, J., Jugessur, A., & Denault, W. R. P. (2023). Stability selection enhances feature selection and enables accurate prediction of gestational age using only five DNA methylation sites. *Clinical Epigenetics*, *15*(1). https://doi.org/10.1186/s13148-023-01528-3

Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, *37*(2), 788–832. https://doi.org/10.1007/s10618-022-00894-5

Kariya, T., Kurata, H., & Hayashi, T. (2024). A modelling framework for regression with collinearity. *Journal of Statistical Planning and Inference*, *228*, 95–115. https://doi.org/10.1016/j.jspi.2023.07.001

Koukaras, P., Mustapha, A., Mystakidis, A., & Tjortjis, C. (2024). Optimizing Building Short-Term Load Forecasting: A Comparative Analysis of Machine Learning Models. *Energies*, *17*(6). https://doi.org/10.3390/en17061450

Shirato, K., Oba, K., Matsuyama, Y., & Hagiwara, Y. (2024). Association of longitudinal pet ownership with wheezing in 3-year-old children using the distributed lag model: the Japan Environment and Children's Study. *Environmental Health: A Global Access Science Source*, *23*(1). https://doi.org/10.1186/s12940-024-01087-x

Sztepanacz, J. L., & Houle, D. (2024). Regularized regression can improve estimates of multivariate selection in the face of multicollinearity and limited data. *Evolution Letters*, *8*(3), 361–373. https://doi.org/10.1093/evlett/qrad064

Tay, J. K., Narasimhan, B., & Hastie, T. (2023). Elastic Net Regularization Paths for All Generalized Linear Models. *Journal of Statistical Software*, *106*. https://doi.org/10.18637/jss.v106.i01

Uniejewski, B. (2024). Regularization for electricity price forecasting. *Operations Research and Decisions*, *34*(3), 267–286. https://doi.org/10.37190/ord240314

Wang, W., & Ruf, J. (2022). A note on spurious model selection. *Quantitative Finance, 22*(10), 1797–1800. https://doi.org/10.1080/14697688.2022.2097120