IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2026) 16 (2): 268-277

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN:2222-2995 E-ISSN:3079-3521

University of Kirkuk Journal For Administrative and Economic Science

Shafeeq Balsam Mustafa. The Use of Support Vector Regression as a Filtering Tool for Applying Simple Linear Regression. *University of Kirkuk Journal For Administrative and Economic Science* **(2026) 16 (2):268-277.**

The Use of Support Vector Regression as a Filtering Tool for Applying Simple Linear Regression

Balsam Mustafa Shafeeq

Information Technology department/Technical College of Management- Baghdad/Middle Technical University, Baghdad, Iraq

balsammustafa95@mtu.edu.iq

Abstract: This study investigates Support Vector Regression (SVR) as a pre-processing filter for Ordinary Least Squares (OLS) in a simple linear regression using simulated data at six sample sizes (n = 50, 100, 200, 400, 700, 1000). An ε-SVR with a radial basis function kernel is first fitted; observations identified as support vectors are retained, and OLS is then re-estimated on this subset. Model quality is evaluated by information criteria. For all n, retention ratios remain consistently high (0.81–0.96) while AIC and BIC consistently decrease post-filtering (ΔAIC from −4.26 at n=50 to −313.29 at n=1000, with almost identical ΔBIC). This implies a meaningful increase in penalized likelihoods without a change to OLS parameterization. Thus, these results support SVR as a reliable denoising filter for enhancing a linearized signal in future estimation, creating a low-cost and easily repeatable avenue to improve classical regression results when faced with noisy or structurally mismatched observations.

Keywords: Support Vector Regression, Simple Linear Regression, Simulation, Kernel Function.

استخدام انحدار متجهات الدعم كأداة ترشيح لتطبيق الانحدار الخطي البسيط

م.د. بلسم مصطفى شفيق

قسم تكنولوجيا المعلومات/الكلية التقنية الإدارية – بغداد/الجامعة التقنية الوسطى، بغداد، العراق

balsammustafa95@mtu.edu.iq

المستخلص: تبحث هذه الدراسة في استخدام انحدار متجهات الدعم (SVR) كمرشح تمهيدي قبل تطبيق طريقة المربعات الصغرى الاعتيادية (OLS) في نموذج الانحدار الخطي البسيط، وذلك باستخدام بيانات مُحاكاة عند ستة أحجام عينات (n = 50، ١٠٠، ٢٠٠، ٤٠٠، ٧٠٠، ١٠٠٠). تم أولاً تقدير نموذج ε-SVR بإستخدام دالة نواة الأساس الشعاعي (RBF)، ثم الاحتفاظ بالمشاهدات التي تم تحديدها كمتجهات دعم، وبعد ذلك أُعيد تقدير نموذج OLS على هذه المجموعة الفرعية. تم تقييم جودة النموذج باستخدام معايير المعلومات. أظهرت النتائج أنه عبر جميع أحجام العينات، بقيت نسب الاحتفاظ مرتفعة بشكل ثابت (٠,٨١–٠,٩٦)، في حين انخفضت قيم معياري AIC وBIC بصورة منتظمة بعد عملية الترشيح (ΔAIC من –٤,٢٦ عند n=50 إلى –٣١٣,٢٩ عند

n=1000، مع قيم ΔBIC شبه متطابقة). ويشير ذلك إلى تحسن معنوي في دوال الاحتمال المعاقبة دون تغيير في معلمات نموذج OLS. وعليه، تدعم هذه النتائج استخدام SVR كمرشح فعال لإزالة الضوضاء وتعزيز الإشارة الخطية في التقدير المستقبلي، مما يوفر آلية منخفضة التكلفة وسهلة التكرار لتحسين نتائج الانحدار الكلاسيكي عند التعامل مع بيانات مشوشة أو غير متطابقة هيكلياً.

الكلمات المفتاحية: انحدار متجهات الدعم، الانحدار الخطي البسيط، المحاكاة، دالة الكرنال.

Corresponding Author: E-mail: balsammustafa95@mtu.edu.iq

Introduction

Regression analysis is one of the primary tools used in understanding variable relationships and making predictions based on observed information in analyzing and predicting important trends. Two of the most popular techniques used as part of this construction are Support Vector Machine (SVM) Regression and Simple Linear Regression Cortes, C (1995). Although different in their approach and equations grounded in mathematics, both are used commonly with the same goal of discovery within data and using those discoveries to predict with accuracy. Support Vector Regression (SVR) is a powerful supervised learning algorithm that is known for its versatility in processing both linear and non-linear regression tasks Zhang & O'Donnell (2020). It identifies the ideal hyperplane that best fits the data points in a multidimensional space. By employing a kernel function, SVR can transform input data into higher-dimensional feature spaces wherein linear separation becomes feasible. This flexibility to complex data distribution renders SVR particularly good at capturing complicated relationships between variables and, thus, offering strong predictive capabilities Hastie, T, 2009). Conversely, Simple Linear Regression represents a basic technique in statistical modeling, praised for its clarity and understandability Montgomery, D. C (2021). Essentially, Simple Linear Regression seeks to establish a linear relationship between a dependent variable and a single independent variable. By fitting a straight line to the observed data points, Simple Linear Regression seeks to assess the extent to which variations in the independent variable impact the dependent variable. Despite its name, the implications gleaned from Simple Linear Regression can be powerful, providing insights into the directionality and strength of relationships between variables. This paper investigates the nuances of Support Vector Regression and Simple Linear Regression. From theoretical insights to practical applications to comparative advantages, we seek to clarify the rationale behind these techniques and provide illustrative examples that demonstrate their real-world usage. Empowered with a better understanding of how to apply these regression techniques through Supported Vector Regression and Simple Linear Regression, we give readers the tools to confidently analyze their own projects. Whether faced with complex arrays of information or simple deductions that require strategic conclusions, Support Vector Regression and Simple Linear Regression serve as formidable allies in helping make sense out of seemingly incomprehensible data.

1st: Literature Review

Where engineering improvements and security developments stem, one can note a dependence upon increasingly sophisticated strategies to solve increasingly complex problems. For example, in tuning cavity microwave filters, Wu and Luo note that manual tuning is limited by nonlinearity and unclear relationships between the tuning direction. By introducing a parametric model that combines principal component analysis and multi-output least squares fuzzy support vector regression, Wu and Luo apply a chaotic ant colony algorithm to obtain tuning parameters. Their results not only suggest high accuracy but also effective predictability concerning the filter characteristics of an eighth-order cavity, laying the foundation for subsequent real-time tuning efforts and improvements for mass production in microwave engineering. In a similar vein, engineering improvements for security are observed in maritime security with vessel trajectory prediction. According to the findings by Luo et al, the extended Kalman filter outperforms least squares support vector regression and enhanced least squares support vector regression in trajectory accuracy and operational efficiency. Further, Sun et al studied transformer temperature prediction as an indicator of reliability for one of the last links of a stable power supply. Given poor transformer cooling methods over time and challenges with heat dissipation, effectively predicting hotspot temperature is critical. Thus, Sun et al create a particle filter optimization support vector regression model that predicts hotspot temperature based on transformer operating data, successfully validated against actual operating data from a 35 kV transformer. Lastly, Qiang et al focus on the resampling particle filter's support vector regression of increasing robustness for localization problems of non-linear filtering problems that differentiate shortcomings among standard particle filtering methods. By resampling particles based on the posterior probability density function rather than using particle confusion as the predicted state, significant improvements in robustness are noted. Therefore, these works advocate for increasingly complex models to challenge the basic understanding of engineering developments, whether for security or successful application, and prove that advanced findings can help better understand situations and improve processes.

2nd: Methodology

1- Support Vector Regression (SVR)

SVR attempts to reduce the amount of error for test data not yet seen while training and employs constraints to generalize the model. It uses a kernel function to project values into a higher-dimensional space in which a linear relationship exists between the input features and the output, M. Eberts (2013). Thus, it can create a nonlinear regression using a linear equation in that newly transformed space. SVR differs from other regressions because it does not train to fit a predetermined structure of the model to the data but instead learns from the data itself. It has been applied successfully in diverse contexts requiring prediction, like time series forecasting, financial modeling, and complex engineering analysis. SVR has also uncovered neurophysiological networks from brain imaging in disease investigations by revealing distributed activation patterns Xiang, D. H (2012). Kernels, sparsity induced by slack variables, and support vectors controlling the margin of separation between classes define the SVR approach. It offers outstanding predictive accuracy and generalization due to being insensitive to input dimension sizes. However, the choice of kernel, cost of constraint violation, and model complexity hyperparameter settings impact final performance Awad (2015).

2- Kernel support vector regression (KSVR)

Kernel support vector regression (KSVR) predicts and analyzes in many disciplines. Regression models are built using SVM and kernel functions. In stock market index prediction, KSVR performed well in terms of least error analysis Montesinos (2022). It has also estimated surface water quality characteristics including chemical and biochemical oxygen demand. In cloud data centers, support vector regression predicts host utilization utilizing various resource utilization histories more accurately than other methods.

3- ε- support vector regression models

The main objective of ε-SVR is to estimate a function with a maximum divergence of ε from the actual response value Chen (2023). This is achieved by composing an ε-insensitive tube symmetrically bounding the calculated function figure (1).

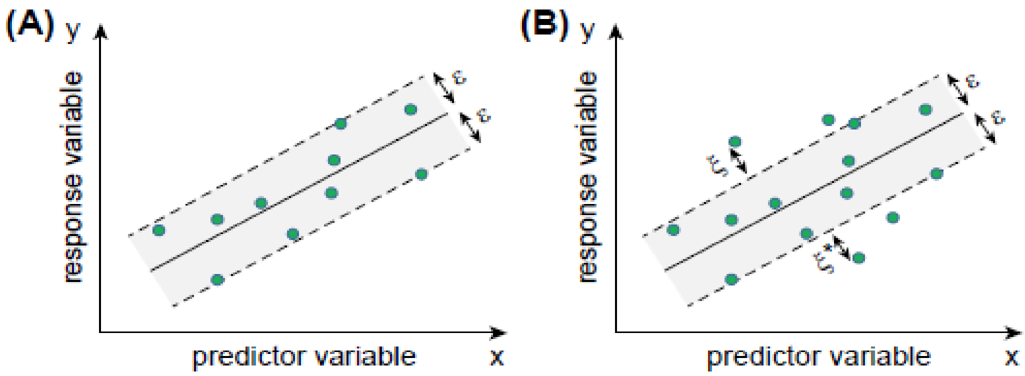

Figure (1): Typical shapes of loss (error) functions

The mathematical formulation of a linear ε-SVR can be expressed as follows:
Let the training data set equal$\{(x_1, y_1), \dots, (x_n, y_n)\}$, where x_i is the input data and y_i is the target output.

The case of a linear function f takes the form:

$$y = f(x)$$
$$y = w, x + b$$
$$y = w^T x + b \qquad (1)$$

Where w, x indicate to the dot product of input data x and the weight vector w.
In ε-SVR, approximation of function f is performed by finding an ε-insensitive tube as flat as possible, which is formally referred to as flatness, i.e., seeking a small w. This can be done by minimizing the norm of w:

$$min \frac{1}{2} \|w\|^2$$

$$\text{Subject to} \begin{cases} y_i - w^T x_i - b \le \varepsilon \\ w^T x_i + b - y_i \le \varepsilon \end{cases} \qquad (2)$$

According to equation (2), ε-SVR involves a linear regression with an ε-insensitive loss function, penalizing guesses outside ε from the desired output. The value of ε determines the flatness of the tube, with a narrow tube indicating low prediction error tolerance and a broad tube indicating high error tolerance. Several ε-insensitive loss functions, such as linear and quadratic functions (3 and 4), are common.

$$L(y, f(x)) = \begin{cases} 0 & , if |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon & , otherwise \end{cases} \qquad (3)$$

$$L(y, f(x)) = \begin{cases} 0 & , if |y - f(x)| \le \varepsilon \\ (|y - f(x)| - \varepsilon)^2 & , otherwise \end{cases} \qquad (4)$$

We can also write the linear loss function with ε-insensitivity zone in equation (2-3) as follows:

$$e(x, y, f) = \max(0, |y - f(x, w)| - \varepsilon) \qquad (5)$$

Thus, the loss is equal to 0 if the difference between the predicted $f(x_i, w)$and the measured value y_i is less than ε.
Note that the ε-insensitivity loss function in equations (3) and (5) defines an ε tube. If the predicted value is within the tube, the loss (error or cost) is zero, and for all other predicted points outside the tube, the loss equals the magnitude of the difference between the predicted value and the radius ε of the tube, figure (2)

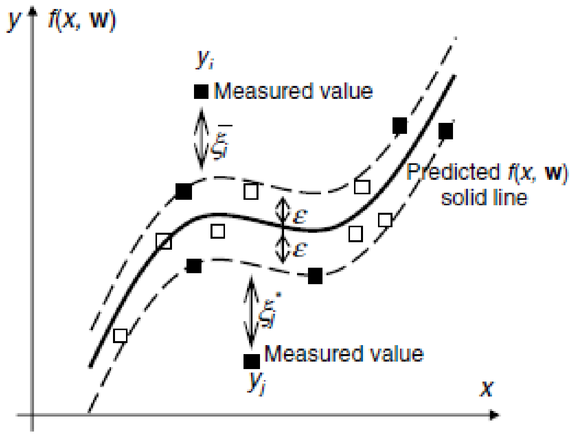

Figure (2): the parameters used in (1-dimensional) SVR

The filled ■ are support vectors, and the empty □ ones are not. Hence, SVs can appear only on the tube boundary or outside the tube.

The original optimization problem in (6) is now written as a multiobjective optimization problem with additional parameters ξ and ξ*

$$min\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

$$\text{Subject to}\begin{cases} y_i - w^T x_i - b \le \varepsilon + \xi_i \\ w^T x_i + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i \ge 0, \xi_i^* \ge 0 \end{cases} \quad (6)$$

Where C > 0 is a regularization parameter, ξ_i and ξ_i^* are slack variables for measurements "above" and "below" an ε-tube, respectively. These variables can account for noisy data at the hyperplane boundaries. Figure (2). The regularization parameter balances approximation error, weight vector norm ‖w‖, prediction error, function flatness, training error, and model VC dimension. Increasing C value penalizes larger training set errors, resulting in a decrease in approximation error Eberts, M (2013). This is achieved by increasing the weight vector norm ‖w‖. A large C value prioritizes prediction errors, ignoring solution complexity, while a small C value prioritizes flatness. In this scenario, the ε-insensitive loss function.

$$L(y, f(x)) = L(\xi) = \begin{cases} 0 & if\ |\xi| \le \varepsilon \\ |\xi| - \varepsilon & otherwise \end{cases} \quad (7)$$

4- Simple Linear Regression

A Simple Linear Regression model is a foundational statistical technique used to understand the relationship between two continuous variables. In essence, it seeks to model the relationship between a dependent variable (often denoted as Y_i) and an independent variable X_i Cardoso (2024). The relationship is represented by a linear equation of the form:

$$Y = \beta_0 + \beta_1 X_i + \epsilon \quad (8)$$

where: β_0 is the intercept.

β_1 is the slope.

ϵ is the error term.

The goal of Simple Linear Regression is to estimate the coefficients β_0 and β_1 based on the available data, minimizing the sum of squared errors (SSE) between the observed values of Y_i and the values predicted by the model. The parameters can be estimated by using the Ordinary Least Squares (OLS) method Hodeghatta (2023):

$$\hat{\beta} = (X^T X) - X^T Y \quad (9)$$

5- The Use of Support Vector Regression as a Filtering Tool

Scholars took great care in their approach, testing support vector regression with different kernels based on root mean squared error to determine the most effective. They used filtering to improve the precision and dependability of regression findings. Filtering with the radial basis function kernel from the support vector regression models resulted in the least amount of error. Therefore, this support vector regression filtered the data before simple linear regression was applied to see how it compared to the original findings. This process of selecting the technique for filtering, as well as preprocessing and refinement, suggests a great deal of concern for improving predictive accuracy and ensuring the strength of later analyses. Ultimately, support vector regression as a filtering mechanism improved predictive accuracy and ensured the strength of later regression analyses. Filtering this way guaranteed accuracy within the structural integrity of the analysis while allowing for comparison with linear regression findings to deeply assess regression results.

3rd: Applications

1- Data Description

Simulated data was used; the dependent variable and independent variable are generated via Python software based on a normal distribution with six different sample sizes (50, 100, 200, 400, 700, and 1000).

2- Results and Discussions

Support Vector Regression (SVR) is an extraordinarily powerful tool in prediction and analysis in diverse fields due to its ability to manage complex relationships between variables. This study investigates the power of SVR as a filtering mechanism to provide the best observations from each sample to undertake a linear regression analysis. The simulated data was of 50, 100, 200, 400, 700, and 1000 in size.

Table (1): shows the performance metrics of statistical accuracy for the varying sample sizes.

n	Sv-count	Retention ratio	OLS before		OLS after		Difference	
			AIC	BIC	AIC	BIC	ΔAIC	ΔBIC
50	48	0.96	158.79	162.61	154.52	158.27	-4.26	-4.35
100	81	0.81	372.70	377.91	319.07	323.86	-53.62	-54.04
200	179	0.90	745.97	752.57	687.23	693.61	-58.74	-58.96
400	367	0.92	1498.84	1506.82	1406.30	1414.11	-92.54	-92.71
700	628	0.90	2585.48	2594.58	2386.86	2395.74	-198.63	-198.84
1000	885	0.89	3649.39	3659.20	3336.10	3345.67	-313.29	-313.54

Applying support vector regression (SVR) as a data-reduction filter before ordinary least squares (OLS) yields systematic improvements in penalized likelihood criteria across all examined sample sizes (n = 50, 100, 200, 400, 700, 1000). Retention rates remain high (0.81–0.96; support vectors 48–885), yet post-filter OLS exhibits uniformly lower AIC/BIC relative to pre-filter OLS within each n: ΔAIC ranges from −4.26 (n=50) to −313.29 (n=1000) with near-identical ΔBIC (−4.35 to −313.54), and the magnitude of improvement increases monotonically with n. Because the OLS parameterization remains unchanged (intercept and slope), these information-criterion reductions primarily reflect substantial gains in log-likelihood on the retained subset, indicating that the SVR step effectively excises observations that degrade the linear fit (noisy or structurally discordant points).
The figures below demonstrate the collected figures for each sample size, respectively for n = 50, 100, 200, 400, 700, and 1000.

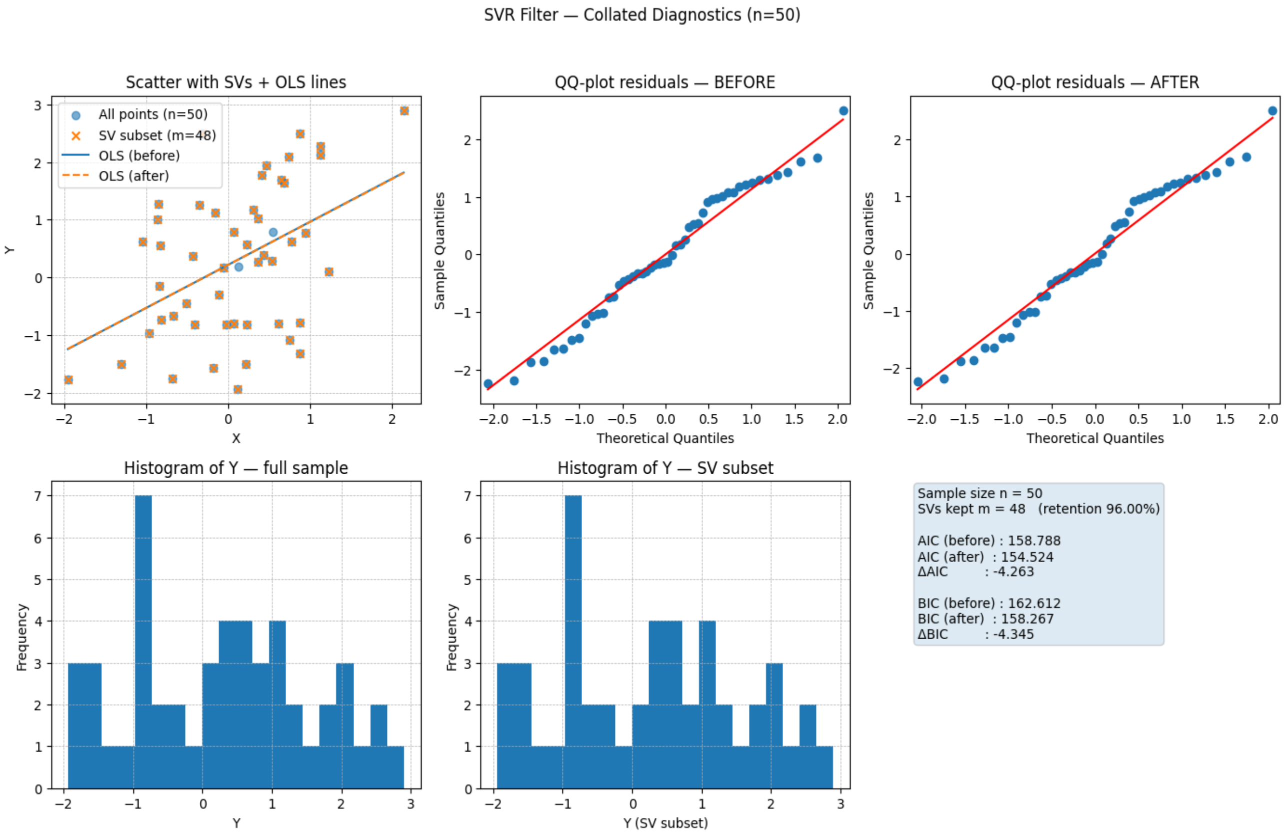

Figure (3)

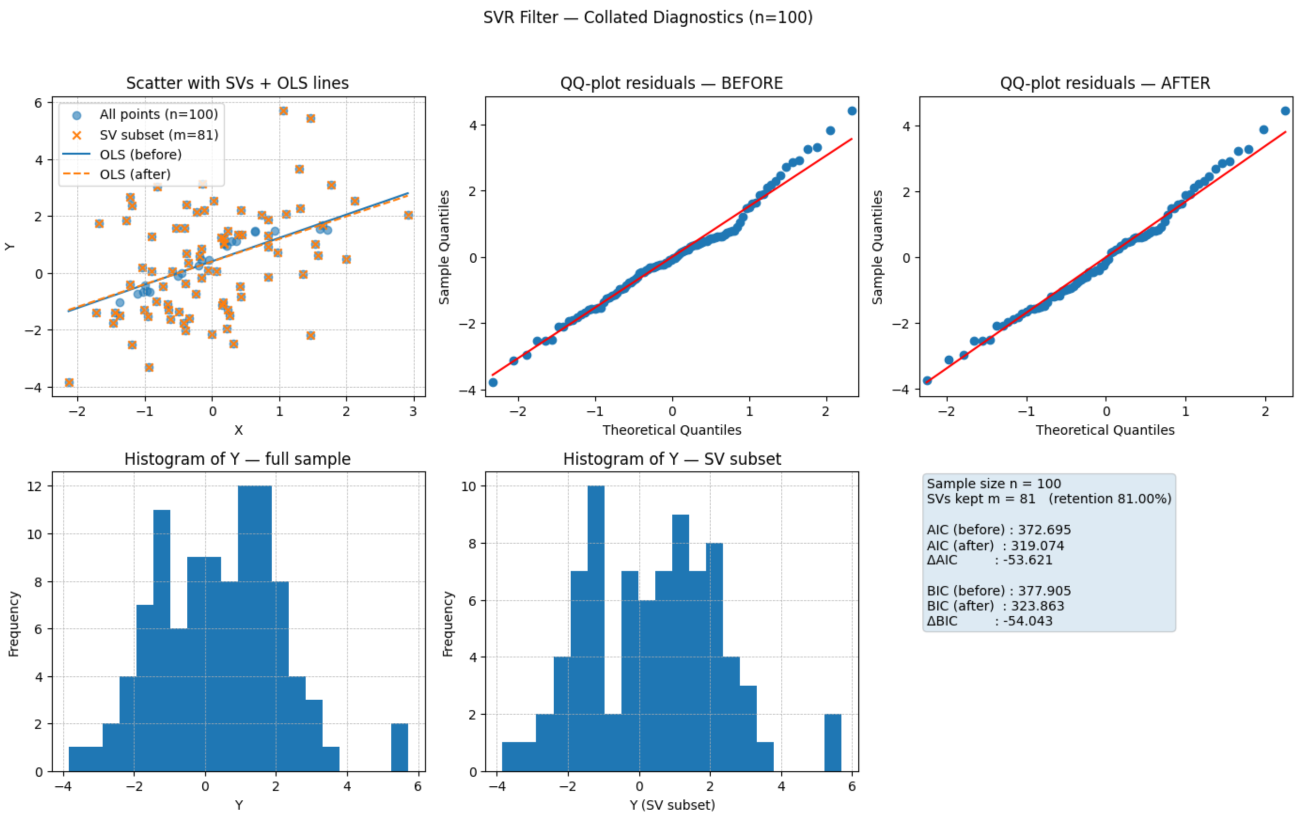

Figure (4)

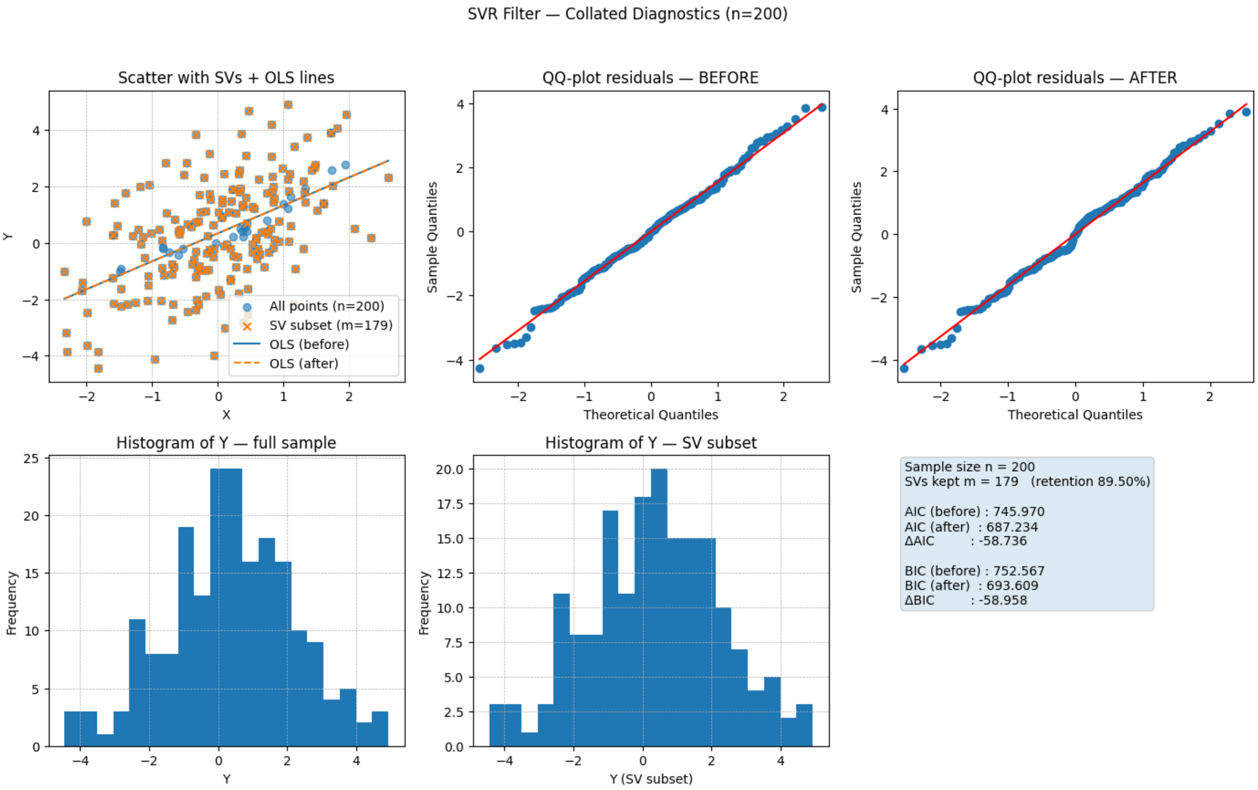

Figure (5)

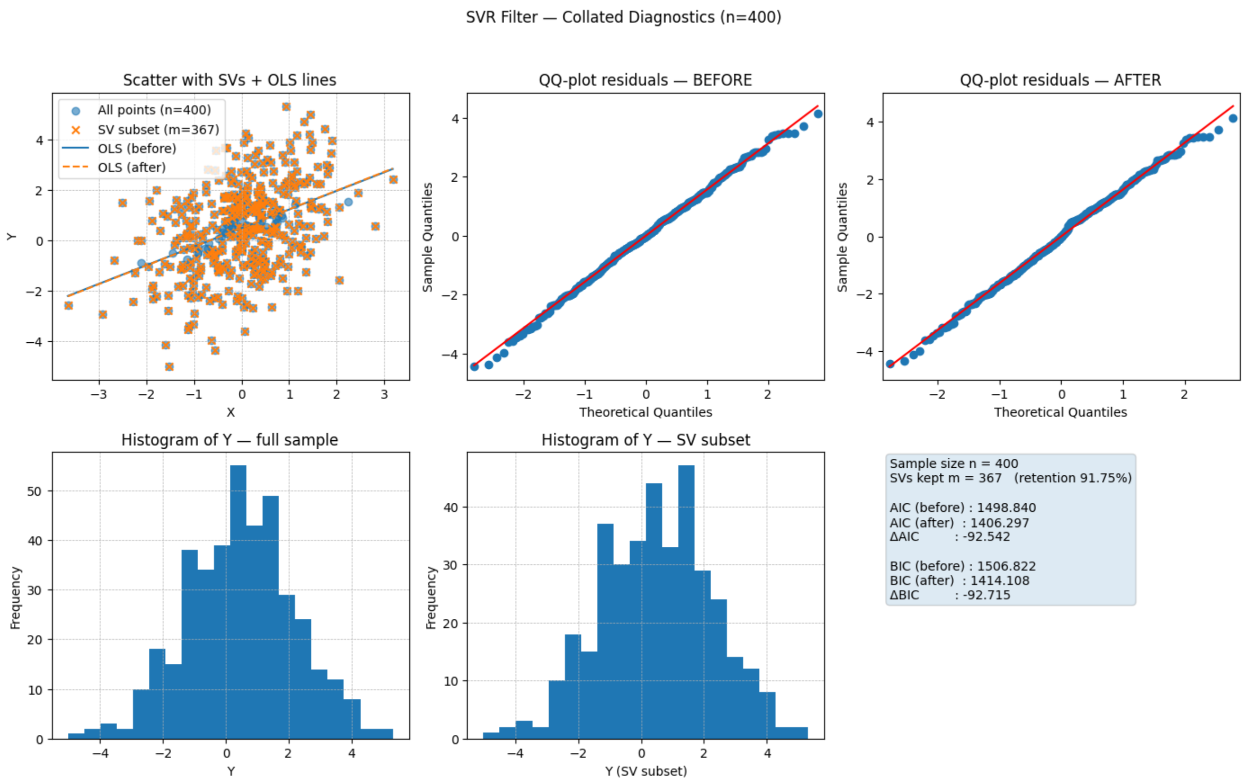

Figure (6)

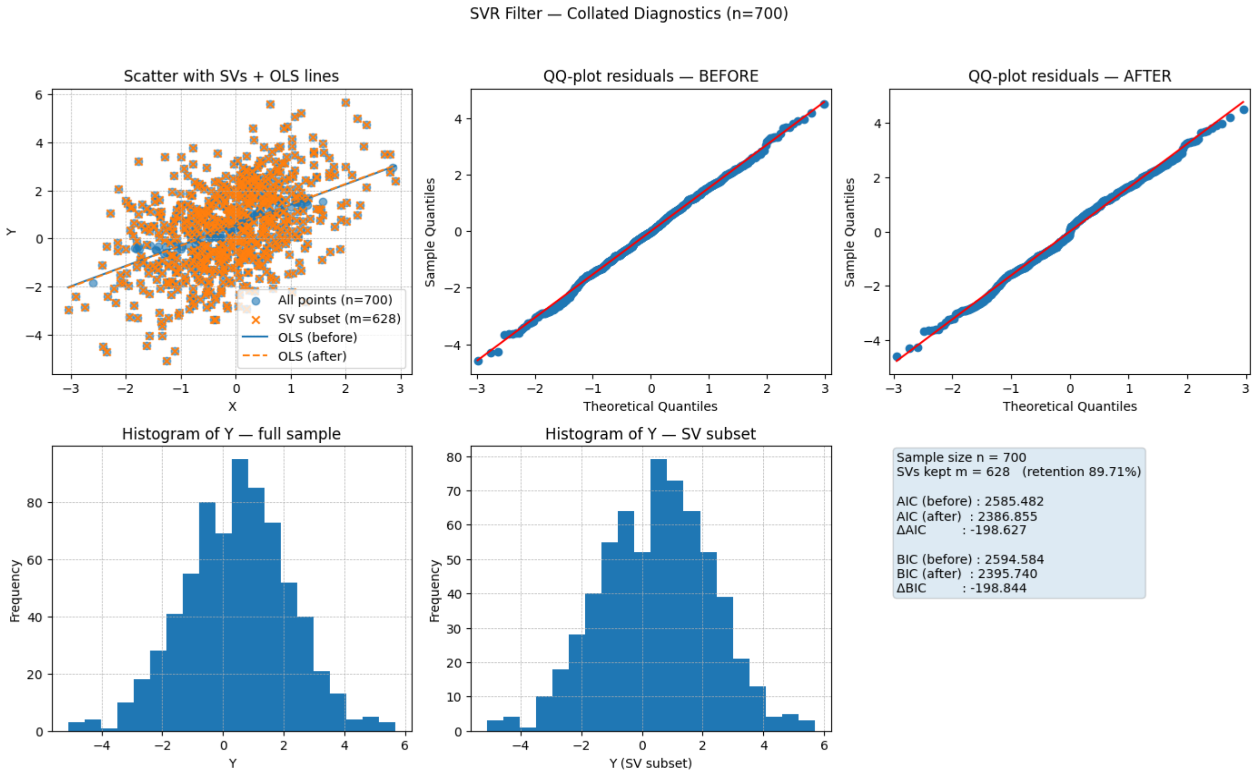

Figure (7)

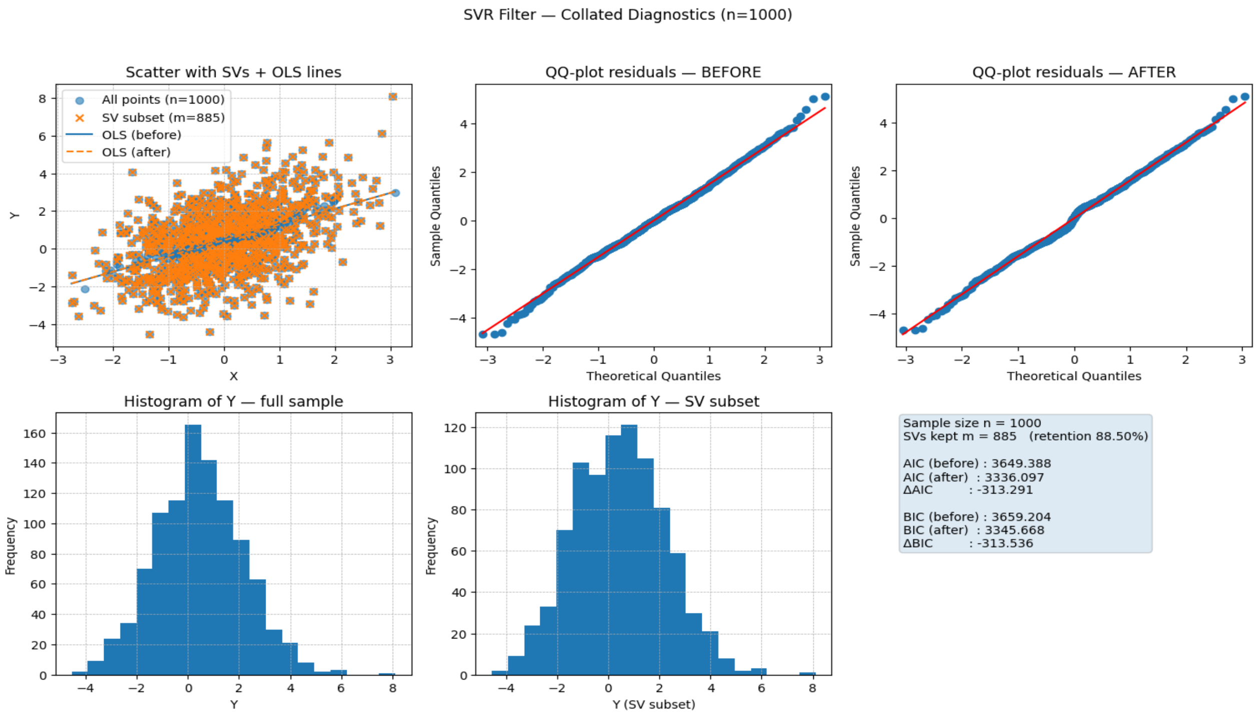

Figure (8)

4th: Conclusion

Employing SVR as a data-reduction filter prior to OLS yields systematic, within-sample improvements in AIC and BIC across all examined sample sizes, with larger samples exhibiting greater absolute reductions. Because the post-filter OLS uses the same linear specification (intercept and slope) as the baseline, the consistent information-criterion gains are attributable to improved fit on the retained subset rather than model complexity. Practically, the approach preserves the interpretability and simplicity of OLS while leveraging SVR's capacity to screen observations that are least compatible with a linear structure. The empirical pattern high retention coupled with marked AICBIC reductions supports the conclusion that SVR filtering can be a pragmatic front-end for enhancing linear modeling in the presence of noise or mild misspecification.

5th: Limitations

The analysis relies on simulated univariate designs with a single predictor and assumes a specific SVR configuration (ε, C, kernel) and filtering rule (retain all support vectors), which may not be optimal or universally applicable. Information-criterion comparisons are made within fixed n; because the filtered sample is smaller, direct cross-n comparisons are not meaningful, and the improvements reflect in-sample penalized likelihood rather than out-of-sample generalization. Moreover, results pertain to homoskedastic Gaussian errors for OLS (and to the chosen data-generating processes explored elsewhere in the manuscript), leaving robustness under heteroskedasticity, heavy tails, leverage points, or stronger nonlinearities untested.

6th: Future Study

Future work should (i) incorporate out-of-sample evaluation (CV or rolling/hold-out tests) and uncertainty quantification to assess predictive benefits; (ii) examine multivariate settings with multiple regressors, interactions, and heteroskedastic or non-Gaussian disturbances; (iii) compare SVR filtering against alternative robust or sparse screens (e.g., Huber/RANSAC, MM-estimators, leverage-based diagnostics, LASSO-based subset selection.

References

1- Awad, M., & Khanna, R. (2015). Support vector regression. In Efficient learning machines: Theories, concepts, and applications for engineers and system designers (pp. 67-80). Berkeley, CA: Apress.
2- Cardoso, R. A., Oliveira, G. A. B., Almeida, G. M. J., & Araújo, J. A. (2024). A simple linear regression strategy for fretting fatigue life estimates. Tribology International, 198, 109852.
3- Chen, K., Liao, Q., Liu, K., Yang, Y., Gao, G., & Wu, G. (2023). Capacity degradation prediction of lithium-ion battery based on artificial bee colony and multi-kernel support vector regression. Journal of Energy Storage, 72, 108160.
4- Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20, 273-297.
5- Eberts, M., & Steinwart, I. (2013). Optimal regression rates for SVMs using Gaussian kernels.
6- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.
7- Hodeghatta, U. R., & Nayak, U. (2023). Simple Linear Regression. In Practical Business Analytics Using R and Python: Solve Business Problems Using a Data-driven Approach (pp. 183-225). Berkeley, CA: Apress.
8- Luo, X., Wang, J., Li, J., Lu, H., Lai, Q., & Zhu, X. (2021, October). Research on Ship Trajectory Prediction Using Extended Kalman Filter and Least-Squares Support Vector Regression Based on AIS Data. In International Conference on Intelligent Transportation Engineering (pp. 1123-1131). Singapore: Springer Nature Singapore.
9- Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Support vector machines and support vector regression. In Multivariate statistical machine learning methods for genomic prediction (pp. 337-378). Cham: Springer International Publishing.
10- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons.
11- Qiang, X., Xue, R., & Zhu, Y. (2021, March). The application of support vector regression in particle filtering. In 2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE) (pp. 5-9). IEEE.
12- Sun, Y., Xu, G., Li, N., Li, K., Liang, Y., Zhong, H., ... & Liu, P. (2021). Hotspot temperature prediction of dry-type transformers based on particle filter optimization with support vector regression. Symmetry, 13(8), 1320.
13- Wu, S., & Luo, X. (2022). A parametric model for cavity filter using multiple outputs least-squares fuzzy support vector regression. International Journal of RF and Microwave Computer-Aided Engineering, 32(10), e23328.
14- Xiang, D. H., Hu, T., & Zhou, D. X. (2012). Approximation analysis of learning algorithms for support vector regression and quantile regression. Journal of Applied Mathematics, 2012(1), 902139.
15- Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. In Machine learning (pp. 123-140). Academic Press.