Canadian AI 2026
25—29 May 2026, Vancouver, BC

Segmentation Expert-Mixture Regularization: An Adaptive Learning Method for Imbalanced Regression Problems

Sarah Elyane Rashiwa[†,*], Paula Branco[†]
[†] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada

Abstract

Imbalanced regression poses a significant challenge for models across a diverse set of domains, where rare extreme cases are often the most important. Standard regression methods, which optimize global error objectives, tend to prioritize high-density regions of the target space, resulting in systematically degraded performance in low-density, extreme regions. Although prior work has focused on data-level strategies that modify the target distribution, comparatively little attention has been devoted to modifying the learning process itself, making it imbalance-aware.

In this paper, we introduce Segmentation Expert-Mixture Regularization (SER), a novel algorithm-level framework for imbalanced regression. SER partitions the target space into regions of varying density and leverages a mixture-of-experts architecture to promote specialization across these regions. A regularization mechanism ensures smooth transitions between the built data partitions and provides a global coherence across segment boundaries. This ensures an adaptive and stable learning method over the entire target space. By integrating segmentation, expert specialization, and regularization within a unified learning framework, SER improves robustness and predictive performance, especially in the rare, extreme, and most important target cases.

Our experiments show consistent improvements over standard models, particularly in extreme target quantiles. We further analyze the impact of segmentation design, parameter sensitivity, and performance variation across the target distribution. To foster reproducibility and future research, all our code is publicly released.

Keywords: Imbalanced Regression, Data Segmentation, Mixture of Experts, Regularization

1. Introduction

Regression under imbalanced target distributions is a pervasive challenge across many real-world applications, including environmental [1, 2], biomedical [3, 4], quality control [5], planning and management [6], e-commerce [7], cybersecurity [8], and finance [9]. In these domains, observations concentrate around central or "normal" operating values. However, the extreme or rare values, although being scarcely represented, often hold the greatest practical importance as they correspond to critical phenomena such as equipment failures, pollution spikes, clinical deterioration, financial crises, or extreme weather events.

Standard regression techniques typically optimize global error objectives (e.g., mean squared error) and therefore prioritize well-represented high-density regions of the target space over less dense regions [10]. As a result, rare extreme outcomes in low-density regions tend to systematically exhibit poor performance. This imbalance leads to increased bias, higher variance, and reduced reliability in the tails of the distribution. Still, the estimation of extreme outcomes is precisely where robust and reliable predictions are most needed, as these rare but high-impact cases often drive decision-making, risk assessment, and intervention strategies in real-world applications.

To address this issue, several prior works—drawing inspiration from the extensive literature on imbalanced classification—have primarily focused on data-level solutions, proposing

[*] srash007@uottawa.ca

This article is © 2026 by author(s) as listed above. The article is licensed under a Creative Commons Attribution (CC BY 4.0) International license (https://creativecommons.org/licenses/by/4.0/legalcode), except where otherwise indicated with respect to particular material included in the article. The article should be attributed to the author(s) identified above.

resampling strategies for regression aimed at rebalancing the target distribution and increasing the representation of rare or extreme values. While such approaches can mitigate distributional skewness, they operate independently of the model's internal learning dynamics and often do not fundamentally alter how the regression function is estimated. In contrast, comparatively less attention has been devoted to adapting the learning process itself to account for target imbalance. Designing regression models that are explicitly tailored to imbalanced target scenarios allows the learning algorithm itself to embed the relative importance of rare and extreme values. This directly influences parameter estimation and model behavior, leading to models that are better aligned with end-user priorities.

In this work, we propose a solution that addresses imbalance directly within the model's learning mechanism. We propose Segmentation Expert-Mixture Regularization (SER), a segmentation-driven approach that partitions the target space into regions of varying density, allowing the model to explicitly distinguish between data regimes of different density. Our expert-mixture strategy encourages specialized predictors to focus on different data regions while maintaining global coherence, and the regularization allows for blending of boundaries. By integrating segmentation with expert specialization and regularization, the proposed framework enables adaptive learning across the entire target space, improving robustness and predictive fidelity in rare but critical regions.

The main contributions of this paper are:

- We propose a novel learning method for imbalanced regression integrating segmentation, mixture of experts, and regularization;
- We carry out experiments to show the performance advantages of our solution;
- We study the impact of our strategy in terms of: the partitions built, the sensitivity of parameters, and the performance variation across target quantiles;
- All our code is freely provided to the research community to allow the reproducibility of our results at https://github.com/srash007/Segmentation-Expert-Mixture-Regularization.

This paper is structured as follows. Section 2 provides a brief literature review on recent progress made in imbalanced regression. In Section 3 we introduce the SER framework, and in Section 4 we carry out the experimental evaluation of our solution and discuss the results obtained. Finally, Section 5 concludes the paper and presents future research directions.

2. Literature Review

Imbalanced regression can be addressed at the data preprocessing level through data-level strategies. These approaches aim to change the data distribution into a more suitable one that allows the learning algorithm to focus on the most relevant and underrepresented target variable cases. Examples of data-level methods include undersampling, oversampling, or a mixture of the two. Several solutions for imbalanced regression problems have been proposed at this level, including methods that generate synthetic samples and methods that work in a distributed fashion (e.g., [11–14]).

Data-level solutions, although easy to apply, pose challenges regarding the amount of change to apply while ensuring that the learning algorithm is correctly biased towards the truly important cases [15]. Algorithm-level approaches, on the other hand, do not suffer from these specific limitations, as they directly modify the learning objective or optimization process rather than altering the data distribution. By embedding imbalance-awareness into the model itself, they enable the regression function to explicitly account for the relative importance of rare and extreme target values. For this reason, they are often regarded as more principled and potentially more effective solutions. However, these methods usually require problem-specific adaptations of the learning algorithm, which can limit their generality and transferability across different model families or application domains.

Even though several algorithm-level solutions exist for imbalanced classification, this remains a poorly explored topic for imbalance regression. Some recent algorithm-level methods have been proposed targeting modifications to the loss function. A typical solution consists of applying modifications to the loss function. For instance, the Label Distribution Smoothing (LDS) [16] and DenseLoss [17] methods aim to approximate the ground truth label density from empirical data using kernel density estimations. The estimated density is then used to change the weights of the data labels. A novel loss function applied is the context of visual imbalanced regression is proposed in the BalancedMSE method [10].

To address imbalanced regression, some researchers have focused on the relationship between feature and target space [16], while other works seek to learn additional tasks to enable the regularization of the model's feature representation [18, 19].

Despite recent advances, algorithm-level strategies for imbalanced regression remain limited in scope and diversity [20], often focusing on loss-function modifications and adapting solutions proposed for class imbalance. Few models explicitly embed imbalance-awareness into their design, revealing the need for algorithm-level approaches that address imbalance directly within the learning process.

3. SER: Segmentation Expert-mixture Regularization for Imbalanced Regression

In this section, we introduce the Segmentation Expert-mixture Regularization (SER) framework, a regression methodology designed for learning under imbalanced regression distributions. The proposed approach is based on the assumption that the functional relationship between predictors and the response variable is not globally homogeneous [21, 22], but varies across different regions of the target space. In such settings, a single global estimator tends to concentrate its modeling capacity in high-density regions, leading to systematic bias on rare and extreme values.

3.1. Notation and SER Overview

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ denote a dataset containing N points, where $\mathbf{x}_i \in \mathbb{R}^d$ represents the d input variables and $y_i \in \mathbb{R}$ is a continuous target variable exhibiting a non-uniform marginal distribution.Let $\mathcal{J}$ be the model selection criterion used for model selection, computed as the Root Mean Squared Error (RMSE) on the training data.The data-generating process is assumed to follow $Y = f(X) + \varepsilon$, where f is an unknown regression function, X represents the input variables, Y the target variable, and ε denotes centered noise that may be heteroscedastic. Empirical diagnostics on real-world imbalanced regression datasets frequently indicate that the behavior of f differs substantially between central and tail regions of the target distribution, which motivates region-specific modeling.

The SER framework addresses this heterogeneity by operating directly in the target space, partitioning it using one defined segmentation strategy. Specialized models are built independently for each one of the partitions by selecting the most suitable from different model families. Finally, regularization is applied in the neighborhood of the transition between two regions to smooth the predictions obtained by local experts. The next section describes the details of the SER Algorithm.

3.2. SER Algorithm

The SER Algorithm, presented in Algorithm 1, uses two thresholds l and u to define a partition of the target variable into lower, central, and upper regions. These thresholds are estimated exclusively from the training data using data-driven criteria, allowing the segmentation to adapt to different imbalance structures without imposing symmetry assumptions in the partitions. Once the partitions are defined, the training set is decomposed into region-specific subsets corresponding to each segment of the target space. We propose four alternative segmentation criteria for this stage: i) Mean Absolute Deviation (MAD)-based; ii) quantiles-based; iii) iterative trimming-based; and iv) bias–variance search–based.

The four criteria are defined as follows:

- **MAD:** defines the two thresholds using the median values and the Mean Absolute Deviation (MAD), providing robustness against extreme values.
- **Quant:** the lower and upper thresholds are defined using fixed percentiles. In particular, we used the 10th and 90th percentiles of the target distribution.
- **Iter:** Thresholds are defined by iteratively removing each tail cases to maximize model performance. Extreme target values on the tails are progressively removed while monitoring global model performance, yielding data-driven thresholds.
- **Bias-Var:** The thresholds are selected by minimizing a bias–variance criterion in the central region, with a penalty for unbalanced segment sizes. This is introduced as an additive term in the optimization and increases when one or more segments contain too few observations. Its role is to prevent degenerate segmentations that would lead to unstable, high-variance expert estimates due to insufficient data support.

Following segmentation, the next stage concerns building specialized models (Algorithm 1 lines 7 to 14). Within each region, an independent regression expert is trained using only observations from the training set assigned to that region. The localized estimation allows each expert to approximate a region-specific regression function with reduced structural bias, while maintaining limited model complexity. Different parametric families may be considered for each region, and the final expert is selected according to predictive performance on the corresponding subset. This localized modeling strategy reflects the assumption that segmentation reduces functional heterogeneity and enables simpler models to perform effectively.

SER does not assume that different target regions require different model families. In practice, the same parametric form may remain appropriate across regions, while the optimal parameters may vary substantially.

However, the fitted coefficients differ across segments, reflecting changes in the relative importance and interactions of predictors across the different extreme-value regimes. A single global estimator averages these incompatible parameter regimes, whereas SER allows the same functional form to be specialized through region-specific parameter estimation.

During inference, the true target value is unavailable and cannot be used to determine region membership. To avoid information leakage, SER relies on a preliminary global estimator trained on the full training set. For a new observation x, a preliminary prediction $\tilde{y}$ is computed and used solely as a gating signal to identify the corresponding region of the target space. Under mild consistency assumptions on the global estimator, the predicted region provides a reliable approximation to the true region, ensuring the validity of the procedure.

Hard assignment to region-specific experts may introduce discontinuities near segmentation boundaries. To mitigate this effect, SER incorporates a regularization mechanism based on α-blending (Algorithm 1 lines 18 to 30). In neighbourhoods surrounding each threshold, predictions are computed as smooth convex combinations of adjacent expert outputs, with interpolation weights determined by the relative position of the preliminary prediction within the transition interval. This formulation ensures continuity of the overall regression function and stabilizes predictions in regions where region membership is uncertain.

The resulting SER predictor defines a single continuous regression function over the entire input domain while preserving region-specific specialization. The framework remains modular and interpretable, and can be applied with different segmentation strategies and expert families, making it suitable for a wide range of imbalanced regression settings.

Algorithm 1 Segmentation Expert-mixture Regularization (SER)

Require: Training data $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, segmentation rule $\mathcal{S}$, blending width $\alpha > 0$, base functional form f_0, set of functional forms $\mathcal{H}$, model selection criteria $\mathcal{J}$

function TRAIN($\mathcal{D}, \mathcal{S}, \mathcal{H}, \mathcal{J}$)
  $\hat{f}_0 \leftarrow$ TRAINMODEL($f_0, \mathcal{D}$) ▷ Train a global regression model
  $(\ell, u) \leftarrow \mathcal{S}(\{y_i\}_{i=1}^n)$ ▷ Compute thresholds
  $\mathcal{D}_L \leftarrow \{(\mathbf{x}_i, y_i) \in \mathcal{D} \mid \hat{f}_0(y_i) \leq l\}$ ▷ Partition the training set into 3 regions
  $\mathcal{D}_C \leftarrow \{(\mathbf{x}_i, y_i) \in \mathcal{D} \mid l < \hat{f}_0(y_i) < u\}$
  $\mathcal{D}_U \leftarrow \{(\mathbf{x}_i, y_i) \in \mathcal{D} \mid \hat{f}_0(y_i) \geq u\}$
  for each functional form $h \in \mathcal{H}$ **do** ▷ Train local experts
    $f_L^h \leftarrow$ TRAINMODEL($h, \mathcal{D}_L$)
    $f_C^h \leftarrow$ TRAINMODEL($h, \mathcal{D}_C$)
    $f_U^h \leftarrow$ TRAINMODEL($h, \mathcal{D}_U$)
  end for
  $f_L \leftarrow \arg\min_{f_L^h} \mathcal{J}(f_L^h)$ ▷ Select best model based on the selection criteria $\mathcal{J}$
  $f_C \leftarrow \arg\min_{f_C^h} \mathcal{J}(f_C^h)$
  $f_U \leftarrow \arg\min_{f_U^h} \mathcal{J}(f_U^h)$
end function

function PREDICT(x, α)
  $\tilde{y} \leftarrow \hat{f}_0(x)$
  if $\tilde{y} \leq \ell - \alpha$ **then**
    return $f_L(x)$
  else if $\ell - \alpha < \tilde{y} < \ell + \alpha$ **then**
    $w \leftarrow (\tilde{y} - (\ell - \alpha))/(2\alpha)$
    return $(1 - w)f_L(x) + wf_C(x)$
  else if $\ell + \alpha \leq \tilde{y} \leq u - \alpha$ **then**
    return $f_C(x)$
  else if $u - \alpha < \tilde{y} < u + \alpha$ **then**
    $w \leftarrow (\tilde{y} - (u - \alpha))/(2\alpha)$
    return $(1 - w)f_C(x) + wf_U(x)$
  else
    return $f_U(x)$
  end if
end function

3.3. Discussion on Segmentation Design and Number of Regions

The use of three target-space segments in SER reflects a common empirical structure observed in imbalanced regression problems. Important and informative observations often lie in the tails of the target distribution, where values are rare, atypical, or extreme, and where predictive behaviors frequently differ from those observed in the central, well-represented region of the data [23, 24].

While the central region is typically characterized by smoother and more homogeneous patterns, tail regions may follow distinct trends that are not adequately captured by a single global model. Moreover, the lower and upper tails do not necessarily share the same functional relationship with the predictors: in practice, the behavior of the lower tail may differ from both the center and the upper tail, and vice versa.

The three-segment formulation should therefore be understood as a flexible and relative decomposition rather than a rigid modeling assumption. It provides a minimal structure capable of capturing asymmetric and heterogeneous behaviors across the target space. When the data do not exhibit meaningful differences between one of the tails and the central region, the effective segmentation naturally reduces to a two-region scenario, without adverse impact on the SER framework.

Conversely, increasing the number of segments beyond three may lead to excessive data fragmentation in low-density regions, amplifying estimation variance and reducing the reliability of local experts. The three-region design thus represents a practical balance between expressiveness and statistical stability.

4. Experimental Evaluation

4.1. Experimental Settings

The experimental evaluation focuses on assessing the empirical performance of SER across different target imbalance conditions and data regimes [15]. In particular, the experiments are designed to examine how target-space segmentation interacts with sample size and tail sparsity, and to identify the settings in which region-specific experts are expected to be most effective. We evaluate SER on several benchmark imbalanced regression datasets [25].

Datasets. We considered three datasets with an imbalanced continuous target variable: *Bank8FM*, *CpuSm*, and *Boston Housing*. *Bank8FM* contains 4,198 observations with eight continuous predictors and presents a strongly right-skewed target distribution with a pronounced upper tail. The *cpuSm* dataset comprises 8,192 instances and exhibits a heterogeneous distribution, with few near-zero values and extreme observations. The *Boston Housing* dataset is included as a low-sample reference case with only 506 cases, serving as a stress test to examine the behavior of SER when data density in the tails is limited and region-specific expert estimation becomes challenging.

Experimental Setup. For all datasets, a random fixed 70/30 train–test split is employed. Experiments are repeated over three different random seeds, which are kept consistent across all methods to ensure fair comparison and reproducibility. All hyperparameter selection and segmentation-related computations are performed exclusively on the training data.

SER is compared against global regression baselines that do not rely on target-space segmentation, namely Ordinary Least Squares (OLS) regression [26] and Huber regression [27], used as representative global estimators. Given the novelty of the proposed segmentation-based learning procedure, we deliberately start with simple baseline models to ensure interpretability and to isolate the contribution of SER before extending the comparison to more complex nonlinear or data-level approaches.

Predictive performance is evaluated on the held-out test sets using the coefficient of determination (R^2) [28], mean absolute error (MAE), and root mean squared error (RMSE) [23]. No resampling, reweighting, or synthetic data generation techniques are applied.

4.2. Results and Discussion

Results overview. Table 1 reports overall performance across datasets. SER consistently improves R^2 and reduces RMSE on the *Bank8FM* and *cpuSm* datasets exhibiting strong imbalance and sufficient sample size. On *Boston Housing*, we couldn't observe gains, which is expected in low-sample regimes where tail regions contain few observations and expert estimation becomes less reliable.

Global estimators fail under target heterogeneity. Figures 1 and 2 show the compute thresholds and the target distributions of *Bank8FM* and *CpuSm* datasets. We observe that

Dataset	Model	R^2	MAE	RMSE
Bank8FM	OLS	0.9321	0.0287	0.0393
	Huber	0.9314	0.0282	0.0395
	SER-MAD	0.9567	0.0241	0.0314
	SER-Quant	**0.9588**	0.0230	**0.0306**
	SER-Iter	0.9460	0.0257	0.0351
	SER-Bias-Var	0.9580	**0.0226**	0.0309
cpuSm	OLS	0.7063	6.1938	9.6092
	Huber	0.0655	13.2588	17.1139
	SER-MAD	**0.9623**	**2.3944**	**3.4295**
	SER-Quant	0.7887	3.2408	7.5940
	SER-Iter	-0.0349	4.1852	14.8127
	SER-Bias-Var	0.9173	3.4851	5.0860
Boston	OLS	**0.6950**	**3.4196**	**4.8652**
	Huber	0.6361	3.6500	5.3166
	SER-MAD	-17.8798	7.7455	31.7394
	SER-Quant	-22.4420	10.6365	40.6047
	SER-Iter	-182.1788	11.6318	73.0549
	SER-Bias-Var	-190.3506	32.0123	111.8212

Table 1. Performance comparison on *Bank8FM*,*cpuSm*, and *Boston* datasets, between regression baselines and SER variants. Results correspond to averages computed over three different random seeds. Best results per dataset and metric are represented in bold.

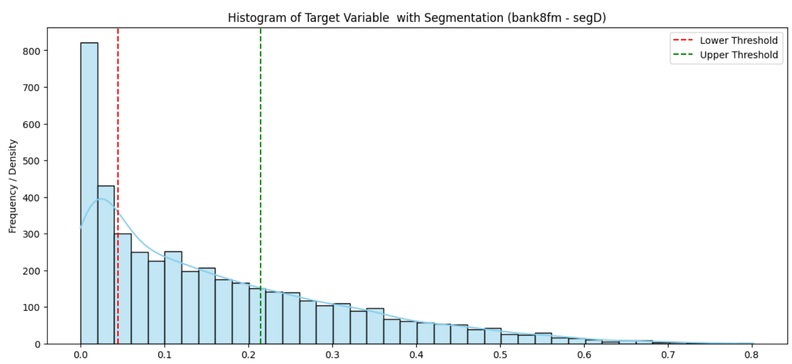


Figure 1. Target variable distribution in *Bank8FM* with data-driven segmentation thresholds selected by the bias–variance criterion. The histogram highlights the strong right-skewness of the target distribution, while the dashed vertical lines indicate the lower and upper thresholds defining the central target region used by SER.

these dataset exhibit sparse regions with highly asymmetric densities [29]. To further illustrate the impact of target heterogeneity, Figures 3a and 4a show observed versus predicted values for a single global OLS model and SER with different segmentation strategies applied. SER improves tail behavior while preserving central performance. Figures 3 and 4 illustrate the effect of target-space segmentation. Compared to the global model, SER yields tighter alignment across heterogeneous target regions, particularly in the tails.

In an ideal setting, points would concentrate around the diagonal line. On *cpuSm*, the scatter reveals strong deviations from this ideal behavior: predictions for very small target values exhibit large variance and include extreme mispredictions (including negative values),

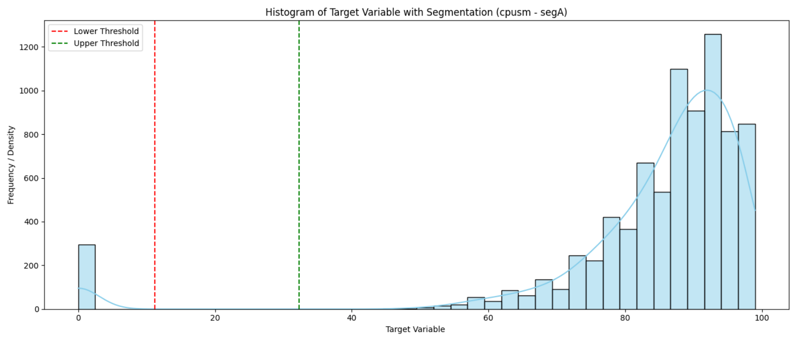


Figure 2. Target variable distribution in *cpuSm* dataset with segmentation thresholds selected by the MAD-based criterion. The histogram reveals a strong concentration of observations at high target values and a sparse lower tail, while the dashed vertical lines indicate the lower and upper thresholds defining the central target region used by SER.

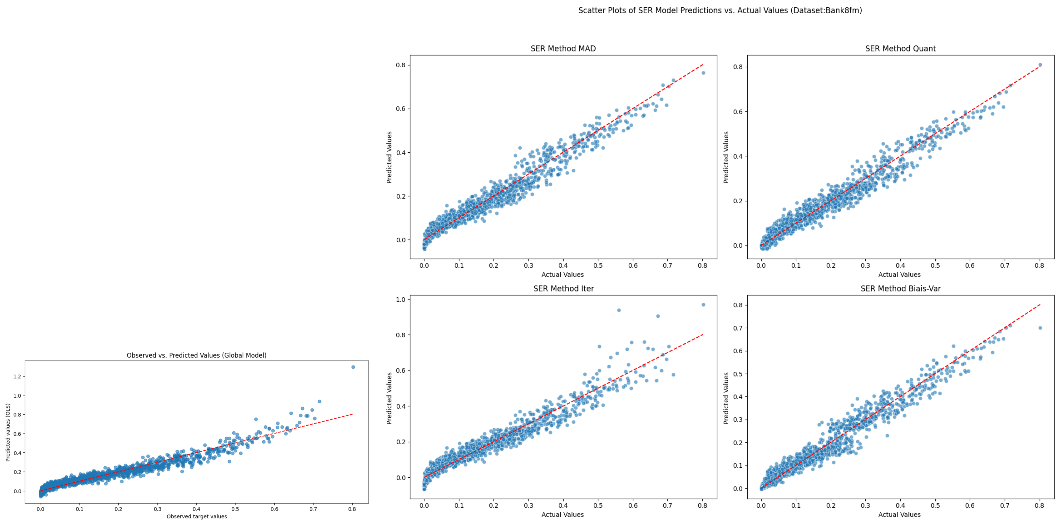
while higher targets show noticeable dispersion around the diagonal. This pattern is consistent with severe target imbalance and heterogeneity: the global estimator is dominated by high-density regions and fails to represent other regions, leading to unstable tail behavior.

Figures 3b and 4b show the corresponding scatter plots for SER under the four segmentation strategies. Compared to the global model, SER yields tighter alignment with the diagonal across the main target range, indicating reduced bias and improved calibration in heterogeneous regions. On *cpuSm*, Methods based on dispersion-aware thresholds (MAD) and the bias–variance objective produce the most stable scatter patterns, with a visibly reduced spread around $y = \hat{y}$ in the dominant high-value region while avoiding the extreme mispredictions observed with the global model. In contrast, quantile-based and iterative strategies display increased dispersion and occasional large outliers, suggesting that segmentation boundaries can become noisy when rare regions have limited support or when thresholds are overly sensitive to sparsity. These qualitative observations align with the quantitative results in Table 1, where SER variants based on MAD and bias–variance segmentation achieve the greatest improvements in R^2 and RMSE on *cpuSm* and *Bank8FM*.

Effect of the segmentation rule (cpuSm case study). Table 1 provides a detailed comparison of segmentation strategies on all datasets, including the *cpuSm*. Performance varies substantially across SER variants, indicating that segmentation quality is a key driver of success. In particular, SER-MAD yields the strongest results across R^2, MAE, and RMSE, suggesting that dispersion-based thresholds can identify meaningful rare, central, and extreme regions under heterogeneous targets. In contrast, purely rank-based segmentation (SER-Quant) and iterative removal (SER-Iter) provide weaker performance, which may reflect sensitivity to sparsity and noisy boundaries in extreme regions with limited support.

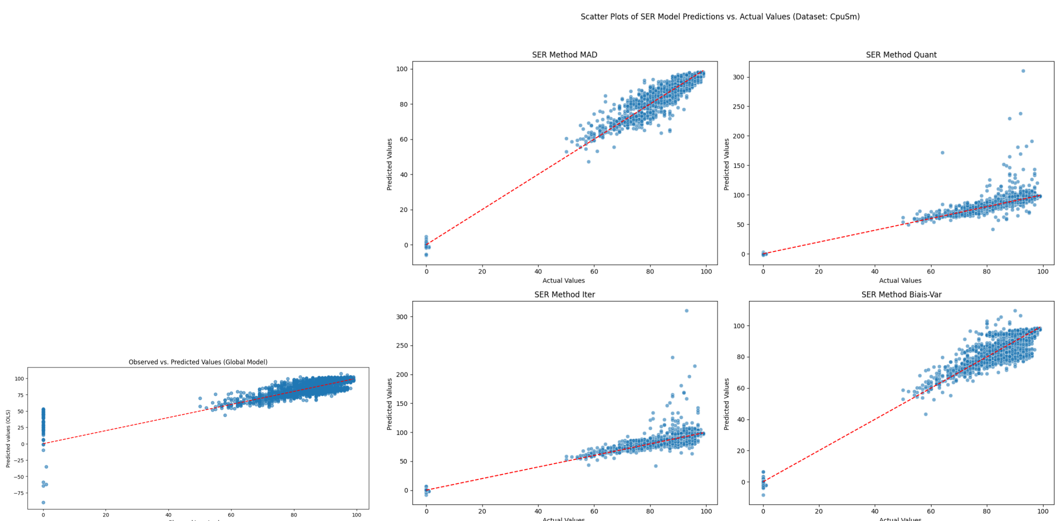
Robust regression is insufficient without structural specialization. Huber regression is designed to reduce the influence of outliers, but it remains a single global estimator. On *cpuSm*, the Huber baseline performs poorly (Table 1), suggesting that the dominant challenge is not only outliers but also a genuinely non-homogeneous mapping across target regions. SER addresses this limitation by allocating dedicated experts to distinct target regimes, rather than applying a uniform robust loss globally.

Low-sample limitation and expected behavior. On *Boston Housing*, improvements are less clear (Table 1 and Figure 5). This behavior is consistent with the fact that tail bins may contain very few samples, limiting the reliability of expert training and segmentation threshold estimation. Boston Housing, therefore, illustrates an expected limitation of segmentation-based approaches: reliable specialization requires a sufficient number of



(a) Global OLS *(b)* SER (four segmentation strategies)

Figure 3. True vs predicted target values on the *Bank8FM* test set. (**Left**) Single global OLS regression. (**Right**) SER with four segmentation strategies. Dashed line represents perfect predictions. SER exhibits tighter alignment, particularly for larger target values, highlighting its ability to mitigate target heterogeneity under strong imbalance.



(a) Global OLS *(b)* SER (four segmentation strategies)

Figure 4. True vs predicted target values on the *cpuSm* test set. (**Left**) Single global OLS regression. (**Right**) SER with four segmentation strategies. Dashed line represents perfect predictions. SER reduces dispersion in heterogeneous target regions; the MAD-based and bias–variance segmentation strategies exhibit the most stable alignment.

observations within each region. Importantly, SER does not catastrophically degrade performance in this regime, indicating that the framework remains stable even when the benefits of specialization are constrained by data scarcity.

Ablation study on sensitivity to parameter α in SER Algorithm. Table 2 shows the results of a set of experiments on *Bank8FM* dataset with varying value of α. We observe

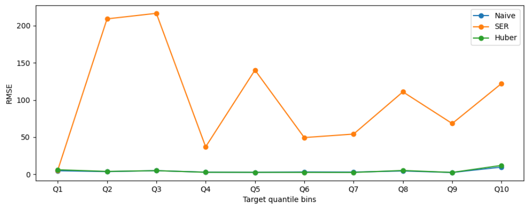


Figure 5. RMSE across target quantile bins for *Boston* dataset, showing the prediction error variation on different target variable regions in a limited-sample setting.

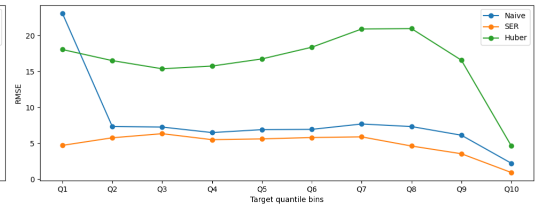


Figure 6. RMSE across target quantile bins for *CpuSm* dataset, showing the prediction error variation on different target variable regions in a limited-sample setting.

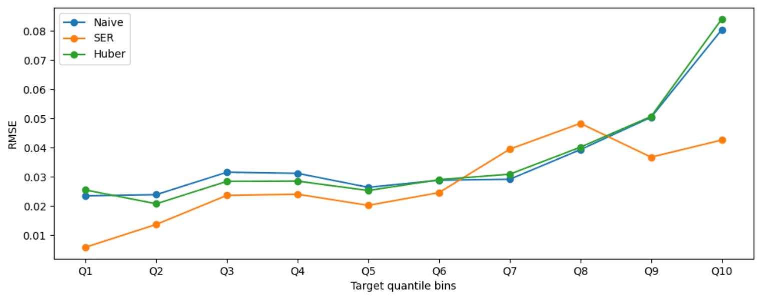


Figure 7. RMSE across target quantile bins for *Bank8FM* dataset, showing the prediction error variation on different target variable regions in a limited-sample setting.

SER	α	R^2	**mae**	**rmse**
MAD	0.01	0.957185	0.024205	0.031848
	0.03	0.957211	0.024207	0.031839
	0.05	0.957287	0.024188	0.031810
	0.07	0.957435	0.024170	0.031755
	0.09	0.957676	0.024136	0.031665
	0.11	0.957896	0.024102	0.031583
	0.13	0.958056	0.024076	0.031523
	0.15	0.958214	0.024046	0.031463
	0.17	0.958434	0.023988	0.031380
	0.19	0.958647	0.023931	0.031300
Quant	0.01	0.959481	0.023159	0.030983
	0.03	0.959600	0.023125	0.030937
	0.05	0.959655	0.023165	0.030916
	0.07	0.959589	0.023225	0.030941
	0.09	0.959579	0.023273	0.030945
	0.11	0.959558	0.023329	0.030953
	0.13	0.959554	0.023374	0.030955
	0.15	0.959569	0.023407	0.030949
	0.17	0.959582	0.023428	0.030944
	0.19	0.959638	0.023419	0.030923

SER	α	R^2	**mae**	**rmse**
Iter	0.01	0.938911	0.027572	0.038043
	0.03	0.939184	0.027601	0.037958
	0.05	0.938817	0.027648	0.038072
	0.07	0.938712	0.027642	0.038105
	0.09	0.938609	0.027626	0.038137
	0.11	0.938801	0.027600	0.038077
	0.13	0.939282	0.027583	0.037927
	0.15	0.939671	0.027598	0.037805
	0.17	0.939876	0.027626	0.037741
	0.19	0.940011	0.027652	0.037699
Bias-Var	0.01	0.958015	0.023150	0.031538
	0.03	0.958547	0.022995	0.031338
	0.05	0.958852	0.022919	0.031222
	0.07	0.959298	0.022802	0.031053
	0.09	0.959710	0.022692	0.030895
	0.11	0.960071	0.022569	0.030756
	0.13	0.960426	0.022444	0.030619
	0.15	0.960742	0.022330	0.030497
	0.17	0.961036	0.022240	0.030383
	0.19	0.961356	0.022153	0.030257

Table 2. Ablation study on SER sensitivity to parameter α on Bank8FM dataset.

that, for all segmentation strategies, the results remain stable and are not heavily affected. This confirms that the regularization step requires minimal tuning, with lower values of α providing a good performance.

Takeaway. Overall, the experimental results confirm that SER provides a principled alternative to purely global regression in imbalanced settings. The method delivers the largest gains when (i) tail regions exhibit predictive behaviors that differ from the central region and (ii) the sample size is sufficient to support reliable region-specific experts, which in our experiments corresponds to datasets containing several thousand observations (approximately 2,500 or more). In lower-sample regimes, SER remains stable but yields more modest improvements, consistent with the expected bias–variance trade-off.

5. Conclusion

We propose the SER Algorithm to address the core challenge of regression under imbalanced target distributions. By applying segmentation of the target variable and introducing an expert-mixture with a regularization component, we successfully partition the target space into Lower, Center, and Upper regions, training specialized local experts for each region, and ensuring prediction continuity through α-blending. Applied to benchmark imbalanced regression datasets SER consistently outperformed the global baselines tested (OLS, Huber) in terms of R^2, MAE, and RMSE, particularly on skewed distributions with rare extremes, demonstrating its ability to reduce bias in tails while maintaining central accuracy. SER provides a robust pipeline with four adaptive segmentation strategies (MAD-based, quantile, iterative removal, and bias-variance optimization), automatic local model selection from diverse candidates (OLS, WLS, GLM, polynomial, Huber, Poisson, quantile), and interpretable region-specific insights that reveal structural differences across target regimes. However, limitations emerged on smaller datasets like Boston, where sparse tails hindered stable local estimation, and certain strategies (e.g., iterative removal) suffered from high variance due to overly small regions.

Regarding future work, we plan to extend SER to be able to handle categorical features alongside numerical predictors, broadening its applicability to real-world domains. Moreover, we will consider refinements to the segmentation process—such as dynamic threshold optimization or density-based clustering—to further enhance robustness, especially in low-sample regimes. Finally, we will also improve SER to be modular and interpretable, so that it provides a foundation for imbalanced regression tasks.

References

[1] M. L. Museru, R. Nazari, A. N. Giglou, K. Opare, and M. Karimi. "Advancing flood damage modeling for coastal Alabama residential properties: A multivariable machine learning approach". In: *Science of the total environment* 907 (2024), p. 167872.

[2] S. Toraskar, A. Khan, M Niranjannaik, A. Singh, and K. Gaurav. "AutoML-Fire: Automated machine-learning approach to predict forest fires". In: *Environmental Modelling & Software* (2025), p. 106578.

[3] A. De Furia, P. Branco, and M. Henderson. "Enhancing TSH-based congenital hypothyroidism screening using machine learning and resampling algorithms". In: *BMC Medical Informatics and Decision Making* 25.1 (2025), p. 449.

[4] S. Babaei Rikan, A. Sorayaie Azar, A. Naemi, J. Bagherzadeh Mohasefi, H. Pirnejad, and U. K. Wiil. "Survival prediction of glioblastoma patients using modern deep learning and machine learning techniques". In: *Scientific Reports* 14.1 (2024), p. 2371.

[5] T. Li, L. Chen, and M. Gan. "Quality control of imbalanced mass spectra from isotopic labeling experiments". In: *BMC bioinformatics* 20.1 (2019), p. 549.

[6] H. Liu, Z. Ma, L. Yang, T. Zhou, R. Xia, Y. Wang, Q. Wen, and L. Sun. "Sadi: A self-adaptive decomposed interpretable framework for electric load forecasting under extreme events". In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023, pp. 1–5.

[7] Q.-Q. He, C. Wu, and Y.-W. Si. "LSTM with particle Swam optimization for sales forecasting". In: *Electronic Commerce Research and Applications* 51 (2022), p. 101118.

[8] A. A. Almazroi and N. Ayub. "Deep learning hybridization for improved malware detection in smart Internet of Things". In: *Scientific reports* 14.1 (2024), p. 7838.

[9] X. Huang, C.-Z. Zhang, and J. Yuan. "Predicting extreme financial risks on imbalanced dataset: A combined kernel FCM and kernel SMOTE based SVM classifier". In: *Computational Economics* 56.1 (2020), pp. 187–216.

[10] J. Ren, M. Zhang, C. Yu, and Z. Liu. "Balanced mse for imbalanced visual regression". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 7926–7935.

[11] L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco. "Smote for regression". In: *Portuguese conference on artificial intelligence*. Springer. 2013, pp. 378–389.

[12] P. Branco, L. Torgo, and R. P. Ribeiro. "SMOGN: a pre-processing approach for imbalanced regression". In: *First international workshop on learning with imbalanced domains: Theory and applications*. PMLR. 2017, pp. 36–50.

[13] P. Branco, L. Torgo, and R. P. Ribeiro. "Pre-processing approaches for imbalanced distributions in regression." In: *Neurocomputing* 343 (2019), pp. 76–99. DOI: https://doi.org/10.1016/j.neucom.2018.11.100.

[14] A. De Furia, A. Kulai, P. Branco, and G.-V. Jourdan. "Distresampler-lite: Light distributed resampler for imbalanced regression problems". In: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer. 2025, pp. 380–391.

[15] P. Branco, L. Torgo, and R. P. Ribeiro. "A Survey of Predictive Modelling on Imbalanced Domains". In: *ACM Computing Surveys* 49.2 (2016), pp. 1–50.

[16] Y. Yang, K. Zha, Y. Chen, H. Wang, and D. Katabi. "Delving into deep imbalanced regression". In: *International conference on machine learning*. PMLR. 2021, pp. 11842–11851.

[17] M. Steininger, K. Kobs, P. Davidson, A. Krause, and A. Hotho. "Density-based weighting for imbalanced regression". In: *Machine Learning* 110.8 (2021), pp. 2187–2211.

[18] Z. Wang and H. Wang. "Variational imbalanced regression: Fair uncertainty quantification via probabilistic smoothing". In: *Advances in Neural Information Processing Systems* 36 (2023), pp. 30429–30452.

[19] Y. Gong, G. Mori, and F. Tung. "RankSim: Ranking Similarity Regularization for Deep Imbalanced Regression". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 7634–7649.

[20] J. G. Avelino, G. D. Cavalcanti, and R. M. Cruz. "Resampling strategies for imbalanced regression: a survey and empirical analysis". In: *Artificial Intelligence Review* 57.4 (2024), p. 82.

[21] M. I. Jordan and R. A. Jacobs. "Hierarchical Mixtures of Experts and the EM Algorithm". In: *Neural Computation* 6.2 (1994), pp. 181–214.

[22] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton. "Adaptive Mixtures of Local Experts". In: *Neural Computation* 3.1 (1991), pp. 79–87.

[23] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. 2nd. Springer, 2009.

[24] R. Koenker. *Quantile Regression*. Cambridge University Press, 2005.

[25] P. Branco, L. Torgo, and R. Ribeiro. *Imbalanced Regression Data Sets*. https://github.com/paobranco/Imbalanced-Regression-DataSets. 2018.

[26] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis*. 5th. Wiley, 2012.

[27] P. J. Huber. "Robust Estimation of a Location Parameter". In: *The Annals of Mathematical Statistics* 35.1 (1964), pp. 73–101.

[28] N. R. Draper and H. Smith. *Applied Regression Analysis*. 3rd. Wiley, 1998.

[29] R. J. Carroll. "A Note on Estimating Heteroscedastic Regression Models". In: *Journal of the American Statistical Association* 77.380 (1982), pp. 771–775.