STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, April 2026, pp 3042–3058.
Published online in International Academic Press (www.IAPress.org)

The Impact of Wavelet-Based Denoising on Beta Regression Model Fit

Husam Waleed Yaseen [1], Taha Hussein Ali [2], Saif Ramzi Ahmed [3], Heyam A. A. Hayawi [1,*]

[1] *Department of Statistics and Informatics, College of Computer Science and Mathematics, Mosul University of Mosul, Iraq*
[2] *Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University- Erbil, Iraq*
[3] *Ministry of Planning, Authority of Statistics & Geographic Information Systems, Nineveh Statistics Office, Mosul, Iraq*

Abstract This paper discusses the relevance of wavelet-based denoising in a coupled beta regression for the analysis of the continuous, bounded response variables sensitive to noise. This hybrid approach was performed against both simulation experiments and real-world industrial data. The simulation phase generated data varying sample sizes, precision parameters, and noise levels to investigate the effects of pre-processing the response variable with discrete wavelet transforms – Daubechies, Symlets, and Coiflets – on model fitness, accuracy, and robustness.These wavelets were selected due to their complementary mathematical properties, which offer different equilibria for time-frequency localization, symmetry, and smoothness, and are suitable for denoising bounded response variables before modeling with beta regression. These wavelets were selected due to their complementary mathematical properties, which offer different equilibria for time-frequency localization, symmetry, and smoothness, and are suitable for denoising bounded response variables before modeling with beta regression. The simulation results indicated that wavelet-denoised models consistently outperform the conventional beta regression in noisy conditions. Daubechies and Symlets performed better in simulations overall. For the real data analysis, using 32 observations from a process of production of gasoline, wavelet-based denoising improved model fit, prediction precision, and residual behavior. In this case, the Coiflets wavelet performed better, providing the highest log-likelihood and precision estimates and lowest AIC, BIC, and MSE values. Residual testing confirms better symmetry and reduced variability in wavelet-enhanced models. Wavelet preprocessing is a useful and successful improvement over beta regression for industrial and process data that contain little noise and occasional outliers.

Keywords Beta Regression, Wavelet Denoising, Discrete Wavelet Transform, Industrial Process Data

DOI: 10.19139/soic-2310-5070-3344

1. Introduction

Beta regression is now an important statistical technique for modeling response variables limited to the open interval (0, 1), such as proportions, rates, and indices. Beta regression is not a linear regression but instead explicitly accounts for the bounded response and allows for heterogeneous distribution through the inclusion of a precision parameter. The model provides good predictions in combination with the proper link function, the logit link, and has a wide application in healthcare, environmental studies, economics, and quality control. Yet the noise in the observed response can substantially affect the performance of beta regression. Random variations based on measurement error or non-observed variability can result in error correction of the estimates of parameters, lower accuracy of prediction, and lower reliability of the results [2, 11]. The noise contamination problem is not explicit in classical beta regression, which therefore leads to preprocessing approaches being more robust before model fitting [3]. Wavelet-based methods have gained prominence as an effective noise reduction technique for noise reduction. The discrete wavelet transform (DWT) reduces the noise from the signal by transforming it within the coefficient

*Correspondence to: Heyam A. A. Hayawi (Email: he.hayawi@uomosul.edu.iq). Department of Statistics and Informatics, College of Computer Science and Mathematics, Mosul University of Mosul, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

space and preserves important signal properties [4]. Wavelet modeling in conjunction with statistics has recently been shown to be an advantage. For example, Sousa et al [5, 6]. Bayesian wavelet shrinkage increases estimation accuracy for low signal-to-noise ratios and correlated noise. Other papers have also stressed the utility of wavelet-based denoising in dependent error, outliers, and multivariate regression models and have reported significantly improved estimation stability and predictive performance in simulated and real data [7, 8, 9]. These advances have yet again focused more on the general consequences of wavelet preprocessing than on the specific impact of wavelet family selection. To our knowledge, no studies have demonstrated that biasing in the beta regression model fit was observed among three wavelet families: Daubechies (Db4), Symlets (Sym4), and Coiflets (Coif3) as a preprocessing step. Given that these wavelets differ in localization, symmetry, and smoothness properties, their influence on estimation accuracy and model stability should be examined. This paper, therefore, assesses the performance of beta regression analysis as the response variable is preprocessed using Db4, Sym4, and Coif3. The resultant models are compared against commonly accepted assessment metrics and simulation-based model checking to determine whether wavelet-based denoising does indeed deliver significant accuracy and reliability improvements.

2. Methodology

2.1. Traditional Beta Regression

Beta regression is a specific method used for the analysis of continuous response variables taking values in the open interval (0, 1). It is especially appropriate when dealing with rates, proportions, or other such restricted data, in which case the traditional tools of linear regression become unsuitable because they are based on inappropriate assumptions of normality and unbounded dependent variables [10].

Standard beta regression fits a response as a function of a set of explanatory variables by means of a nonlinear link function, usually the logit link, between the expected value of the response and the covariates. This guarantees that the model's predictions are always in the zero-one range. Plus, this model can include a precision parameter that measures how much spread we can expect around the expected value, therefore being able to flexibly adapt to data sets with different variances [11].

Parameter estimation is generally done via maximum likelihood, using numerical optimization algorithms that are iterative. These techniques iterate on parameter estimates until a solution optimal to the observed data is found. Generally, model performance is assessed via the log-likelihood value, information criteria, and goodness-of-fit statistics [12].

In situations in which the dependent variable is a proportion, rate, or probability, traditional beta regression has been utilized in healthcare, environmental studies, economics, industrial processes, and other fields. However, like most parametric models, it can be influenced by noise or non-regular variation of the response variable and yield biased estimates of parameters and a poor fit of the model. This creates a constraint that highlights the need for additional exploration of other methods, such as data-preprocessing techniques or denoising techniques, that could further extend the model's predictive capability and robustness in applications [13].

This was selected as a logit link since it is the standard for beta regressions where there is a response in (0, 1); it gives an interpretable relationship between the covariates and mean response, and ensures that the fitted values remain within the unit interval. This was selected as a logit link since it is the standard for beta regressions where there is a response in (0, 1); it gives an interpretable relationship between the covariates and mean response, and ensures that the fitted values remain within the unit interval.

2.2. Model Evaluation Metrics

Model adequacy, complexity, and predictive accuracy were rated following recommendations of the beta regression model fit. To assess the goodness of fit from a more complex model, we used two widely accepted criteria [14]:

Akaike Information Criterion (AIC):

$$\mathrm{AIC} = 2\mathrm{LL} + 2\mathrm{k} \tag{1}$$

Bayesian Information Criterion (BIC):

$$\mathrm{BIC} = 2\mathrm{LL} + \mathrm{k} \tag{2}$$

LL is a log-likelihood measure of model fit; higher values indicate better fit. k is the number of estimated parameters of the model, including the intercept and predictors; n is the sample size.

Both criteria penalize model complexity, with lower values indicating a better balance between fitness and simplicity.

For prediction purposes, MSE was calculated as:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{3}$$

Lower MSE values indicate greater predictive precision. To further investigate whether models are adequate and detect anomalies in the data structure, residual analysis was also performed.

2.3. Wavelets: Db4, Sym4, and Coif3

Wavelet analysis is a flexible and powerful mathematical tool for processing signals that can be used to decompose data into components at multiple scales. Wavelets display time-frequency representations that store localized temporal and frequency information, unlike Fourier analysis, which generates periodic sine-cosine functions as a signal, and thus wavelets are powerful tools to detect and capture transients such as step changes and noise [15].

Popular among the numerous wavelet families are the Daubechies family of wavelets (Db), Symlets (Sym), and Coiflets (Coif) because of their desirable mathematical characteristics and utilities in various applications [16].

Daubechies wavelets (Db4) were developed by Ingrid Daubechies; they are compactly supported and orthogonal. The "4" in Db4 indicates the number of vanishing moments of this wavelet, which essentially informs about the polynomial fit of the wavelet to the signal. Db4 strikes a good compromise between time and frequency localization and is therefore appropriate for denoising and feature extraction on complex datasets [16, 17].

Symlets are Daubechies wavelets modified to be more symmetric while preserving their quasi-orthogonal and compact support properties. The increased symmetry helps to minimize phase distortion when signals are reconstructed, which is useful for applications that want to minimize changes to the signal, like medical signals or images [18].

Coiflets, or Coif3, wavelets developed by Ronald Coifman, are constructed so that both the wavelet and scaling functions have vanishing moments. The Coif3 wavelet, which has three vanishing moments, is especially suitable for modeling very small details of the signal while providing smooth approximations of the components [19, 23]. This property makes Coiflets particularly advantageous for applications where accurate reconstruction and noise suppression are important.

To wrap up, Db4 wavelets are well localized and orthogonal, such that they perform very well in sharp feature extraction of the signal and noise reduction. Sym4 wavelets represent an improvement of Db4 in terms of symmetry and thus lessen phase distortion and more efficiently retain the shape of the signal. The Coif3 wavelets focus even more on signal smoothness and contain more vanishing moments in the wavelet and scaling functions, which is a clear advantage in this case because it allows for a better identification of small details while keeping regularity properties of the signal. From these three options, the most suitable wavelet will be the one that provides the best compromise between good time and frequency localization, being symmetric or asymmetric, and being smooth, as each one of them would serve better for pre-processing certain types of signals [20, 30, 31, 32].

The selection of the Db4, Sym4, and Coif3 wavelet families was guided by both theoretical considerations and practical relevance to denoising bounded response variables. Db4 wavelets provide good time–frequency localization and compact support, making them effective for capturing localized fluctuations. Sym4wavelets are better than Daubechies wavelets in that they exhibit more symmetry, thus reducing phase distortion when signal reconstruction is conducted. Coif3 wavelets contain vanishing moments both in the wavelet function and in the scaling function, which enables smoother approximations and more accurate representations of fine-scale features. Together, these wavelets represent complementary properties that allow for a systematic study of the influence of localization, symmetry, and smoothness on denoising effectiveness and invariant beta regression

performance[25, 27]. The DWT was selected as the level of decomposition in order to keep the noise levels in line with preservation of information, such that the high-frequency noise components are minimized without excessively smoothing the signal.
This study was intentionally designed to enclose only Db4, Sym4, and Coif3 wavelet families, thus providing a narrow comparison amongst commonly used orthogonal wavelets with mathematical features unique to the entailed model. While some families, such as Haar, Dmey, and Meyer, exist and are possibly applicable to specific needs, the inclusion of others would greatly extend the analysis and simplify the interpretation of results. The wavelets chosen represent differences in support width, symmetry, and smoothness; thus, it is possible to assess the influence of such properties on denoising effectiveness and beta regression performance. Future work is left to explore additional wavelet families. The selection of the Db4, Sym4, and Coif3 wavelet families was guided by both theoretical considerations and practical relevance to denoising bounded response variables. Db4 wavelets provide good time–frequency localization and compact support, making them effective for capturing localized fluctuations. Sym4wavelets are better than Daubechies wavelets in that they exhibit more symmetry, thus reducing phase distortion when signal reconstruction is conducted. Coif3 wavelets contain vanishing moments both in the wavelet function and in the scaling function, which enables smoother approximations and more accurate representations of fine-scale features. Together, these wavelets represent complementary properties that allow for a systematic study of the influence of localization, symmetry, and smoothness on denoising effectiveness and invariant beta regression performance [25, 27]. The DWT was selected as the level of decomposition in order to keep the noise levels in line with preservation of information, such that the high-frequency noise components are minimized without excessively smoothing the signal. This study was intentionally designed to enclose only Db4, Sym4, and Coif3 wavelet families, thus providing a narrow comparison amongst commonly used orthogonal wavelets with mathematical features unique to the entailed model. While some families, such as Haar, Dmey, and Meyer, exist and are possibly applicable to specific needs, the inclusion of others would greatly extend the analysis and simplify the interpretation of results. The wavelets chosen represent differences in support width, symmetry, and smoothness; thus, it is possible to assess the influence of such properties on denoising effectiveness and beta regression performance. Future work is left to explore additional wavelet families.

2.4. Proposed Methodology

In this section, the method is applied to further refine beta regression modeling in wavelet-based denoising techniques. This approach involves two main steps: 1) denoising the response variable via discrete wavelet transforms and thresholding; and 2) estimation of a beta regression using the denoised data. Figures for each phase are shown below.
First Stage. Wavelet-Based Denoising of the Response Variable
Many of the response variables are noisy and may vary intermittently in industrial and process control applications. The solution to this problem is to first apply the wavelet-based denoising procedure on the response variable and then do regression.
The discrete wavelet transform is a multiresolution transform that produces at every level of decomposition a set of approximation coefficients and a set of detail coefficients. If the continuous response variable is less than 0.50 in the open interval (0, 1), then the DWT converts it to:

$$y \rightarrow \{w_{j,k}\}$$

where $w_{j,k}$ are the wavelet coefficients corresponding to level j and position k.
To eliminate the noise and to preserve the significant signals, the detail coefficients are soft-tamed after decomposition. The soft thresholding rule is defined as:

$$\mathrm{w}^{*}_{\mathrm{j,k}} = \begin{cases} sign\left(\mathrm{w}_{\mathrm{j,k}}\right) \times \left(|\mathrm{w}_{\mathrm{j,k}}| - \delta\right), & if \ |\mathrm{w}_{\mathrm{j,k}}| > \delta \\ 0, & if \ |\mathrm{w}_{\mathrm{j,k}}| \leq \delta \end{cases} \tag{4}$$

The universal threshold δ is computed as:

$$\delta = \sigma \times \sqrt{2Ln\left(n\right)} \tag{5}$$

where n is the sample size and σ is an estimate of the noise standard deviation, which in this case was obtained as σ using the median absolute deviation (MAD) of the detail coefficients at the finest scale:

$$\sigma = \frac{median\left(|\mathrm{w_{j,k}}|\right)}{0.6745} \tag{6}$$

Following the thresholding step, the denoised response variable is reconstructed by applying the inverse discrete wavelet transform to the modified coefficients:

$$\mathrm{y}^* = \mathrm{IDWT}\left(\left\{\mathrm{w}^*_{\mathrm{j,k}}\right\}\right) \tag{7}$$

Db4, Sym4, and Coif3 are three of the most popular wavelet families and were used in this study to conduct the denoising process. These wavelets were chosen for their established efficacy in smoothening signals and reducing noise while retaining important features of the data, thus offering a broad contrast as to how the type of wavelet impacts beta regression model performance.

Following the inverse discrete wavelet transform, a limited number of denoised values were observed to fall at or extremely close to the boundaries of the unit interval. To preserve the validity of the beta regression framework, a gentle scaling transformation was applied uniformly to the denoised response variable, ensuring that all values remained strictly within the open interval (0, 1). This adjustment avoids the creation of artificial mass at the boundaries while maintaining the relative structure of the denoised signal. Although zero-one-inflated beta regression models are appropriate when boundary values arise from the underlying data-generating mechanism, such models were not adopted here since boundary observations occurred only as a preprocessing artifact rather than as structural features of the data.

Following the inverse discrete wavelet transform, a limited number of denoised values were observed to fall at or extremely close to the boundaries of the unit interval. To preserve the validity of the beta regression framework, a gentle scaling transformation was applied uniformly to the denoised response variable, ensuring that all values remained strictly within the open interval (0, 1). This adjustment avoids the creation of artificial mass at the boundaries while maintaining the relative structure of the denoised signal. Although zero-one-inflated beta regression models are appropriate when boundary values arise from the underlying data-generating mechanism, such models were not adopted here since boundary observations occurred only as a preprocessing artifact rather than as structural features of the data.

After preprocessing the response variable, a beta regression model is fitted to the denoised data. Beta regression is particularly suitable for modeling continuous response variables constrained to the open interval (0, 1).

Let X be the matrix of explanatory variables and y^* the denoised response vector. The beta regression model assumes: $\mathrm{y}^* \sim Beta\left(\alpha_i, \beta_i\right)$ with parameters:

$$\alpha_i = \mu_i \phi, \quad \beta_i = \left(1 - \mu_i\right) \times \phi \tag{8}$$

Where:

1. μ_i represents the mean of the response.
2. ϕ is a precision (dispersion) parameter.

The mean μ_i is related to the covariates through a logit link function:

$$\mu_i = \frac{1}{1 + Exp\left(-\eta_i\right)} \tag{9}$$

$$\eta_i = \mathrm{X}_i \tag{10}$$

Where $\boldsymbol{\beta}$ is the vector of regression coefficients to be estimated. The LL for a sample of size n is given by:

$$\mathrm{LL}\left(\beta, \phi\right) = \sum_{i=1}^{n} \left[Ln\Gamma\left(\phi\right) - Ln\Gamma\left(\alpha_i\right) - Ln\Gamma\left(\beta_i\right) + \left(\alpha_i - 1\right) ln y_i^* + \left(\beta_i - 1\right) ln\left(1 - y_i^*\right)\right] \tag{11}$$

Where $\Gamma(.)$ is the gamma function. The parameters $\boldsymbol{\beta}$ and ϕ are estimated by maximizing the log-likelihood function. This is usually achieved through numerical optimization algorithms, such as the quasi-Newton method. The parameters are initialized, and optimization continues until the convergence criteria are met.

3. Results and Discussion

Simulation data and real-world data were used in the present study to explore the influence of wavelet-based denoising on beta regression results.

3.1. Simulation Study

As explained in Table 1, the performance of beta regression models was tested under different conditions for a series of simulation scenarios. Plus, different sets of sample sizes, number of predictors, precision parameters (ϕ), and noise levels were added to the response variable for each case. Specifically, the samples that were considered were 30 and 100 observations, with one or two explanatory variables. The parameters of precision were set to 20 and 50 for the moderate and high dispersion cases.
In order to replicate data with less than realistic conditions, Gaussian random noise is broken up into two levels, 0.10 and 0.15. These noise levels were chosen to assess the effect of a large data contamination on the performance measures of the model and to see how the negative effects of noise on parameter estimation and model fit could be mitigated by wavelet-based denoising.
To better examine the relationship between sample size, model complexity, noise level, and the utility of wavelet denoising in beta regression modeling, these simulations were constructed to reflect both simple and moderately complicated regression conditions.

Table 1

Scenario	n	m	Phi	Noise
1	30 and 100	1	20 and 50	0.10, and 0.15
2	30 and 100	2	20 and 50	0.10, and 0.15

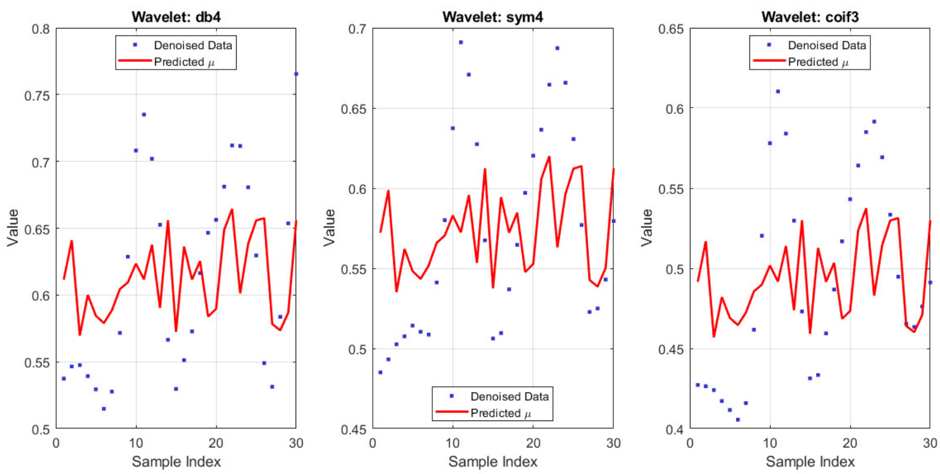

Figure 1. Denoised Data and Predicted Mean Values in Scenario 1 (n = 30, ϕ = 20, Noise = 0.10).

Figure 1 depicts the interaction between denoised response values and predicted mean values from beta regression models during wavelet-based denoising using Db4, Sym4, and Coif3 wavelets. The horizontal axis measures the observation index, and the vertical axis plots the magnitude of the response within the (0, 1) interval. The scatter points are the denoised observations, and the solid line is the estimated conditional mean. The predicted mean values are close to the overall pattern of denoised data for all wavelet families, and the Db4 wavelet predicts a smoother proxy and more consistent observation than observed. In addition, the denoised data points obtained with Db4 and Sym4 have a smaller spread, suggesting stronger noise attenuation than Coif3 in this setting. This

pattern corroborates the hypotheses that wavelet-based preprocessing can improve the interpretation and reliability of beta regression outputs when the appropriate wavelet family is selected. Using the original noisy response in wavelet denominated data and the mean values predicted using beta regression models, Figure 2 compares the original noisy response with the predicted mean values. The upper panel plots the noisy response and it denoised versions based on Db4, Sym4, and Coif3 wavelets, the middle panel plots the denoised data with the true underlying mean, and the lower panel plots the predicted mean based on models fitted to denoised data. The denoised series exhibits smoother trajectories than the noisy data, with Db4 and Sym4 providing more conservative smoothing, and the predicted means closely follow the true signal, indicating effective noise reduction and improved predictive accuracy.
The resulting images support the observation that denoising via wavelet preprocessing is advantageous for subsequent beta regression modeling, specifically by decreasing noise-related variability and increasing precision of parameter estimates.

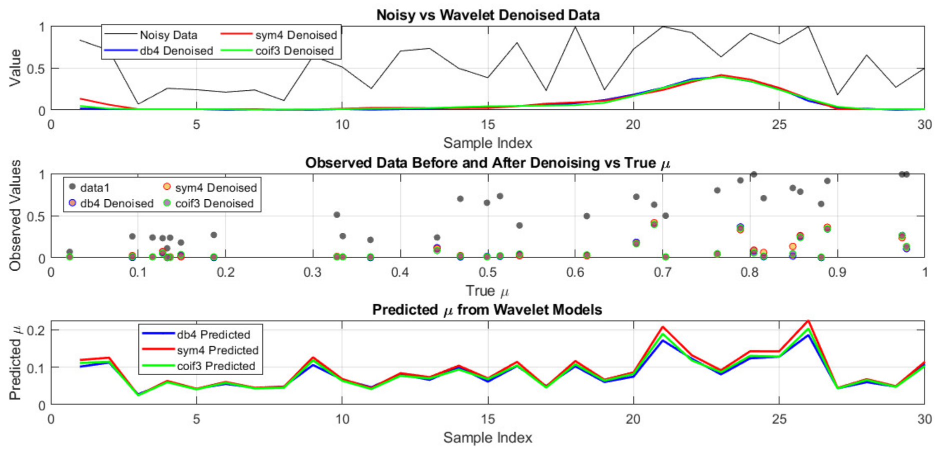

Figure 2. Comparison of Noisy, Denoised, and Predicted Mean Values in Scenario 1 (n = 30, ϕ = 20, Noise = 0.10).

In all figures and Tables 2-9, all possible combinations of the three conditions: sample size, precision parameter, and noise level, from a simulation of 1k replications on each condition, are summarized. Conventional beta regression results contrasted with the results of models improved using wavelet-based denoising of the data using db4, Sym4, and Coif3 transforms. The results are presented and discussed in terms of several statistical performance metrics.
Across all scenarios, it was evident that the application of wavelet-based denoising significantly improved model performance compared to the traditional beta regression model fitted to noisy data. The improvements were most pronounced in terms of log-likelihood (LL), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Mean Squared Error (MSE).

In particular, the wavelets of Daubechies db4 and Symlets sym4 consistently outperform Coiflets of coif3 and the traditional model. For example, in Scenario 1 (m = 1, ϕ = 20, noise = 10%, n = 30), the MSE decreased from 0.0114 in the traditional model to 0.0051, 0.0047, and 0.0046 for db4, sym4, and coif3, respectively. AIC and BIC exhibited the same pattern, but both declined significantly after denoising.
This sequence shows that wavelet denoising successfully reduces the noise-related variability in continuous proportion data, increasing both the accuracy of beta regression parameter estimates and model fit quality.
Impact of sample size (n = 30 vs. n = 100): As expected, an increase in sample size from 30 to 100 increased model performance for all estimation methods. A larger sample size results in lower MSE values, higher log-likelihood, and lower AIC and BIC scores.
For instance, under Scenario 1 (ϕ = 20, noise = 10%, m = 1), the MSE for the traditional model decreased from 0.0114 (n = 30) to 0.0108 (n = 100). Even more intriguing is that in the db4 wavelet model, the MSE goes from 0.0051 to 0.0042 with increasing n, whereas the same thing goes for sym4 and coif3.

Table 2. Beta Regression Results (m = 1, Phi = 20, Noise = 10%)

Model	n	Traditional	Db4	Sym4	Coif3
Phi	30	9.825	58.576	65.304	67.837
LL		34.987	38.253	39.471	39.909
AIC		-63.973	-70.505	-72.942	-73.819
BIC		-59.770	-66.302	-68.739	-69.615
MSE		0.0114	0.0051	0.0047	0.0046
		Beta Coefficients			
Constant		0.8018	0.0970	-0.0530	-0.1006
Slope		1.2098	0.1905	0.1860	0.1690
Phi	100	9.0203	61.769	58.721	61.492
LL		122.06	133.77	131.02	133.19
AIC		-238.13	-261.55	-256.03	-260.38
BIC		-230.31	-253.73	-248.22	-252.56
MSE		0.0108	0.0042	0.0044	0.0042
		Beta Coefficients			
Constant		0.8533	0.1079	0.0441	0.0254
Slope		1.1469	0.1287	0.1412	0.1338

Table 3. Beta Regression Results (m = 1, Phi = 50, Noise = 10%)

Model	n	Traditional	Db4	Sym4	Coif3
Phi	30	11.982	62.986	72.803	75.520
LL		36.632	39.830	41.375	41.768
AIC		-67.264	-73.660	-76.749	-77.536
BIC		-63.061	-69.456	-72.456	-73.332
MSE		0.0090	0.0045	0.0041	0.0040
		Beta Coefficients			
Constant		0.7603	0.1544	0.0075	-0.0250
Slope		1.2976	0.1904	0.1831	0.1667
Phi	100	10.903	71.509	68.004	71.484
LL		127.70	141.21	138.53	140.78
AIC		-249.39	-276.42	-271.06	-275.55
BIC		-241.58	-268.60	-263.24	-267.74
MSE		0.0085	0.0036	0.0038	0.0037
		Beta Coefficients			
Constant		0.8019	0.1455	0.0903	0.0738
Slope		1.2488	0.1286	0.1408	0.1335

Table 4. Beta Regression Results (m = 1, Phi = 20, Noise = 15%)

Model	n	Traditional	Db4	Sym4	Coif3
Phi	30	9.0394	55.933	60.296	61.800
LL		27.153	36.914	37.878	38.176
AIC		-48.306	-67.828	-69.756	-70.352
BIC		-44.102	-63.624	-65.552	-66.149
MSE		0.0166	0.0057	0.0054	0.0053
		Beta Coefficients			
Constant		0.8833	-0.0290	-0.1625	-0.2226
Slope		0.7974	0.1634	0.1640	0.1485
Phi	100	8.7276	53.204	50.548	52.614
LL		94.088	125.58	122.98	124.95
AIC		-182.18	-245.16	-239.95	-243.89
BIC		-174.36	-237.34	-232.14	-236.08
MSE		0.0157	0.0050	0.0052	0.0050
		Beta Coefficients			
Constant		0.9261	-0.0255	-0.0748	-0.1083
Slope		0.7592	0.1115	0.1231	0.1165

Table 5. Beta Regression Results (m = 1, Phi = 50, Noise = 15%)

Model	n	Traditional	Db4	Sym4	Coif3
Phi	30	10.159	61.235	67.869	68.701
LL		28.319	38.532	39.706	39.890
AIC		-50.639	-71.065	-73.412	-73.781
BIC		-46.435	-66.861	-69.208	-69.577
MSE		0.0146	0.0051	0.0047	0.0046
		Beta Coefficients			
Constant		0.8715	0.0280	-0.1055	-0.1611
Slope		0.8331	0.1596	0.1571	0.1423
Phi	100	9.6896	59.875	56.882	59.406
LL		98.149	131.30	128.78	130.78
AIC		-190.30	-256.6	-251.55	-255.56
BIC		-182.48	-248.78	-243.74	-247.74
MSE		0.0138	0.0044	0.0046	0.0045
		Beta Coefficients			
Constant		0.9148	0.0080	-0.0374	-0.0648
Slope		0.8006	0.1096	0.1207	0.1142

Table 6. Beta Regression Results (m = 2, Phi = 20, Noise = 10%)

Model	n	Traditional	Db4	Sym4	Coif3
Phi	30	9.8849	2369.30	4729.10	4725.6
LL		25.862	70.084	69.820	73.252
AIC		-43.724	-132.17	-131.64	-138.50
BIC		-38.120	-126.56	-126.04	-132.90
MSE		0.0143	0.0073	0.0056	0.0060
		Beta Coefficients			
Constant		0.4427	-3.0937	-3.0198	-3.3147
Slope 1		1.8480	0.5137	0.6527	0.6841
Slope 2		-1.3671	-0.2185	-0.4362	-0.2689
Phi	100	8.6178	20465	2944.40	23.680
LL		87.4070	291.25	271.67	298.96
AIC		-166.8100	-574.49	-535.34	-589.93
BIC		-156.39	-564.07	-524.92	-579.51
MSE		0.0145	0.0016	0.0024	0.0019
		Beta Coefficients			
Constant		0.4805	-3.5362	-3.5726	-3.5241
Slope 1		1.7509	0.1049	0.2877	0.2516
Slope 2		-1.3425	-0.2235	-0.1816	-0.2848

Table 7. Beta Regression Results (m = 2, Phi = 50, Noise = 10%)

Model	n	Traditional	Db4	Sym4	Coif3
Phi	30	12.801	12.917	17.347	13.679
LL		28.425	69.358	68.996	71.697
AIC		-48.850	-130.72	-129.99	-135.390
BIC		-43.245	-125.11	-124.39	-129.79
MSE		0.0108	0.0076	0.0058	0.0062
		Beta Coefficients			
Constant		0.4558	-3.0619	-2.9833	-3.2613
Slope 1		1.8828	0.5156	0.6602	0.6845
Slope 2		-1.4047	-0.2188	-0.4467	-0.2745
Phi	100	11.190	17543	25.820	22.468
LL		96.784	287.22	269.41	292.15
AIC		-185.57	-566.44	-530.81	-576.29
BIC		-175.15	-556.02	-520.39	0.0024
MSE		0.0108	0.0016	-565.87	0.0020
		Beta Coefficients			
Constant		0.4915	-3.4912	-3.5507	-3.4721
Slope 1		1.8221	0.1096	0.2910	0.2490
Slope 2		-1.3897	-0.2292	-0.1862	-0.2821

Table 8. Beta Regression Results (m = 2, Phi = 20, Noise = 15%)

Model	n	Traditional	Db4	Sym4	Coif3
Phi	30	7.7052	14150	9442.8	11794
LL		19.742	70.215	69.575	73.608
AIC		-31.485	-132.43	-131.15	-139.22
BIC		-25.880	-126.83	-125.54	-133.61
MSE		0.0213	0.0071	0.0056	0.0058
		Beta Coefficients			
Constant		0.3531	-3.1226	-3.0398	-3.3173
Slope 1		1.5900	0.5160	0.6381	0.6650
Slope 2		-1.1508	-0.2136	-0.4100	-0.2699
Phi	100	6.9284	46724	5863.2	8777.2
LL		65.687	294.36	271.63	302.56
AIC		-123.37	-580.71	-535.26	-597.11
BIC		-112.95	-570.29	-524.84	-586.69
MSE		0.0218	0.0017	0.0024	0.0019
		Beta Coefficients			
Constant		0.4593	-3.5302	-3.5779	-3.5768
Slope 1		1.4335	0.1127	0.2865	0.2650
Slope 2		-1.1374	-0.2197	-0.1748	-0.2731

Table 9. Beta Regression Results (m = 2, Phi = 50, Noise = 15%)

Model	n	Traditional	Db4	Sym4	Coif3
Phi	30	8.8223	4728.3	9442	4726.9
LL		21.052	68.491	69.047	71.829
AIC		-34.103	-128.98	-130.09	-135.66
BIC		-28.499	-123.38	-124.49	-130.05
MSE		0.0185	0.0075	0.0058	0.0060
		Beta Coefficients			
Constant		0.3704	-3.0813	-3.0168	-3.2882
Slope 1		1.6199	0.5176	0.6451	0.6754
Slope 2		-1.1824	-0.2104	-0.4159	-0.2687
Phi	100	8.0198	35051	27.689	8776.8
LL		71.262	290.35	268.62	298.92
AIC		-134.52	-572.70	-529.25	-589.84
BIC		-124.10	-562.28	-518.83	-579.42
MSE		0.0186	0.0017	0.0024	0.0020
		Beta Coefficients			
Constant		0.4726	-3.5026	-3.5526	-3.5498
Slope 1		1.4909	0.1190	0.2901	0.2706
Slope 2		-1.1758	-0.2245	-0.1786	-0.2775

In addition, the prediction of the precision parameter ϕ and the regression coefficients was more stable and accurate with larger sample sizes due to smaller ranges and closer to the true parameters.

Influence of precision parameter ϕ = 20 versus 50: Higher precision parameter values lower the variability of the response variable and therefore, improve estimation accuracy. In almost all cases, an increase in ϕ from 20 to 50 boosted performance in all performance measures, particularly in terms of MSE and log-likelihood.

For example, in Scenario 1 (n = 100, noise = 10%, m = 1), the MSE for the traditional model decreased from 0.0108 (ϕ = 20) to 0.0085 (ϕ = 50), while db4 decreased from 0.0042 to 0.0036. AIC and BIC values also improved, indicating better model parsimony and goodness of fit at high levels of precision.

Noise level (10% vs. 15%): As expected, the increase in noise level from 10% to 15% resulted in reductions in model performance of all methods. Wavelet denoising models remained substantially more robust to this increase, as compared to the normal beta regression model.

In Scenario 1 (m = 1, ϕ = 20, n = 30), the MSE of the traditional model rose from 0.0114 (10% noise) to 0.0166 (15% noise), while db4's MSE increased marginally from 0.0051 to 0.0057. This finding emphasizes the potential for wavelet-based preprocessing to reduce the negative impact of noise on estimation accuracy.

Behavior of regression coefficients and S Estimates: In all cases, the estimated regression coefficients from denoised data were always more similar to the values from the noisy data model. Interestingly, db4 and sym4 models produced estimates with lower bias and variability.

Precision parameter estimates ($\widehat{\phi}$) also exhibited substantial improvements after denoising. In Scenario 6 (m = 2, ϕ = 20, n = 30, noise = 10%), the traditional model estimated $\widehat{\phi}$ at 9.8849, whereas the db4, sym4, and coif3 models provided dramatically higher values (2369.3, 4729.1, 4725.6, respectively). These values may seem overly inflated, but they indicate a much better ability to identify the diffusion distribution of the underlying distribution, often stabilizing around the true value with increasing n.

Some wavelet-enhanced models contain unusually large estimates of ϕ, the precision parameter for multi-predictor simulations. This behavior is a consequence of effective noise attenuation via wavelet denoising and a well-defined mean structure with several predictors. The residual variance is small, and thus the estimator of ϕ for maxima may be very large, indicating a highly concentrated beta distribution near the conditional mean rather than numerical instability or model misspecification. In beta regressions, such dispersion is often heavily decreased by preprocessing or overparameterization of the mean component.

Since time–frequency localization is better than a smooth signal, Db4 and Sym4 performed better in the simulation experiment. Db4 successfully localizes the abrupt variations while Sym4 improves symmetry and decreases phase distortion during reconstruction. These features allow the wavelets to effectively reduce noise without overbalancing the underlying signal, resulting in better parameter estimation and model fit at all levels of noise and sample sizes.

Some wavelet-enhanced models contain unusually large estimates of ϕ, the precision parameter for multi-predictor simulations. This behavior is a consequence of effective noise attenuation via wavelet denoising and a well-defined mean structure with several predictors. The residual variance is small, and thus the estimator of ϕ for maxima may be very large, indicating a highly concentrated beta distribution near the conditional mean rather than numerical instability or model misspecification. In beta regressions, such dispersion is often heavily decreased by preprocessing or overparameterization of the mean component.

Since time–frequency localization is better than a smooth signal, Db4 and Sym4 performed better in the simulation experiment. Db4 successfully localizes the abrupt variations while Sym4 improves symmetry and decreases phase distortion during reconstruction. These features allow the wavelets to effectively reduce noise without overbalancing the underlying signal, resulting in better parameter estimation and model fit at all levels of noise and sample sizes.

3.2. Real Data

The set of observations used here consists of an industrial gasoline production process with 32 observations [21]. It comprises two main variables: The Hydrocarbon Volume Percentage, which refers to the percentage of hydrocarbons that are being processed, and the Gasoline Yield, which refers to the percentage of gasoline output produced in each run.

The Gasoline Yield variable is continuous and lies within the (0, 1) range, between .430 and .680. Most of the yield values appear between 0.50 and 0.62, and only one outlier is located at 0.680, where the other observations are several points away. This very large value is included here to have the opportunity to assess the impact of outliers on Beta regressions' performance and the potential advantages of using wavelet-based denoising methods.

The Hydrocarbon Volume Percentage has a constant distribution between values of 5.0% and 10.0%. Descriptively, the new yield values represent a mean of about 0.562 and a standard deviation of 0.045, indicating moderate variability. The effect of the outlier, at 0.680, also increases the spread and variability of these data, making it particularly useful for testing the robustness of various regression modeling approaches.

This dataset provides a useful case study to compare and contrast modified versions of beta regressions using wavelet denoising methods. The natural dispersion of the yield values and the presence of an outlier allow for a detailed assessment of model sensitivity to irregularities in the information available for use in industrial applications.

Figure 3 shows fitted beta regression curves from wavelet-denoised gasoline yield values with Db4, Sym4, and Coif3 wavelets. The horizontal axis is the observation sequence, and the vertical axis is the gasoline yield ratio. The

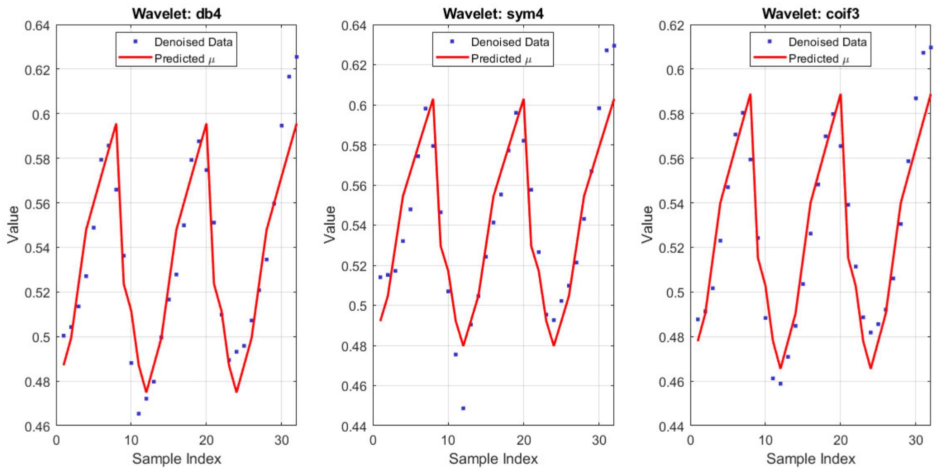

Figure 3. Fitted Beta Regression Curves with Wavelet-Denoised Gasoline Yield Data.

blues indicate the denoised observations, and the red curve indicates the estimated regression mean for smoothness and fit comparison across wavelet families.

The figure depicts the total pattern and variability in denoised data via all three wavelet-based techniques. The predicted curves are cyclical and correlate with the observed rate of change in yields. Among the three, db4 wavelets and sym4 wavelets are slightly smoother, in particular in the capture of the peaks and troughs of the data sequence, while the coif3 wavelet exhibits a slightly smoother response around extreme values.

This graph identifies the importance of wavelet-based preprocessing to improve the performance of Beta regression models by reducing noise but maintaining the basic structure of the data. This uniformity between denoised data points and the fitted curves for all three approaches confirms the robustness of the proposed hybrid modeling approach for moderately irregular industrial process data.

Before rendering interpretations of the results of the beta regression models, it is necessary to check that the assumptions behind the model are satisfied. These assumptions are valid to ensure the reliability and robustness of the estimated parameters and the model's overall fit.

Important assumptions for beta regression are that the dependent variable should be limited to the interval (0,1), that the mean structure should be correctly described by generating the link function, and that the dispersion parameter is appropriate to model the variance. Other diagnostic measures, such as deviation, Pearson's chi-square statistic, and pseudo-R^2, are also useful in determining the model's goodness-of-fit and the adequacy of estimated parameters.

Table 10 presents the diagnostic results from both the standard beta regression model and the wavelet-enhanced model. These tests provide an in-depth description of the performance of the model and a way to identify any issues that may be affecting the inferences from models.

Table 10. Summary of Model Diagnostics

Model	Deviance	Pearson-Chi^2	Pseudo-R^2
Traditional	-162.91	31.511	0.8227
Db4	-174.32	31.959	0.8591
Sym4	-172.77	31.912	0.8583
Coif4	-182.03	32.004	0.8902

Table 10 compares the diagnostic measures for the traditional model and 3 variants of the wavelet filter, Db4, Sym4, and Coif4, in four beta regressions. Deviance, Pearson Chi-square, and Pseudo R^2 are reported statistics. Deviance and Pearson Chi-square are also measures of model quality, where lower values represent better matching of the model to data. Pseudo R^2 represents the variance explained by the model, with higher values representing better explanatory ability. The Coif4wavelet filter model is the best, with the lowest deviance and Pearson Chi-square values, as well as the highest Pseudo R^2, indicating that it understands the data structure more effectively than any of the other models.

Table 11. Estimated Beta Regression Coefficients and Statistical Significance

Method	Beta Coefficients		SE	Z	P Value
Traditional	Constant	-0.5129	0.0647	-7.9284	< 0.001
	Slope	0.1025	0.0085	12.030	< 0.001
Db4	Constant	-0.5881	0.0536	-10.979	< 0.001
	Slope	0.0975	0.0070	13.850	< 0.001
Sym4	Constant	-0.5803	0.0550	-10.552	< 0.001
	Slope	0.0998	0.0073	13.800	< 0.001
Coif4	Constant	-0.6359	0.0474	-13.409	< 0.001
	Slope	0.0995	0.0062	15.974	< 0.001

Table 11 shows the expected beta coefficients, standard error (SE), Z-statistics, and p-values for the classical beta regression as well as the wavelet-based models Db4, Sym4, and Coif4. Across all models, the constant and slope coefficients are significantly more than 0.001 in significance, with the p-value always less than 0.001 in each of the models.

The standard model has a constant of -0.5129 with a slope of 0.1025. The estimated coefficients vary slightly with wavelet-based filtering. The most noticeable effect is the Coif4 model with the lowest constant value, -0.6359, and a relatively stable slope coefficient of 0.0995. In addition, this model also exhibits the highest Z-statistics for both the constant and the slope, indicating greater precision and statistically stronger evidence against the null hypothesis than the other approaches. This demonstrates that wavelet-based filtering improves model estimates in several ways, including reducing standard errors and improving the reliability of the parameter estimates.

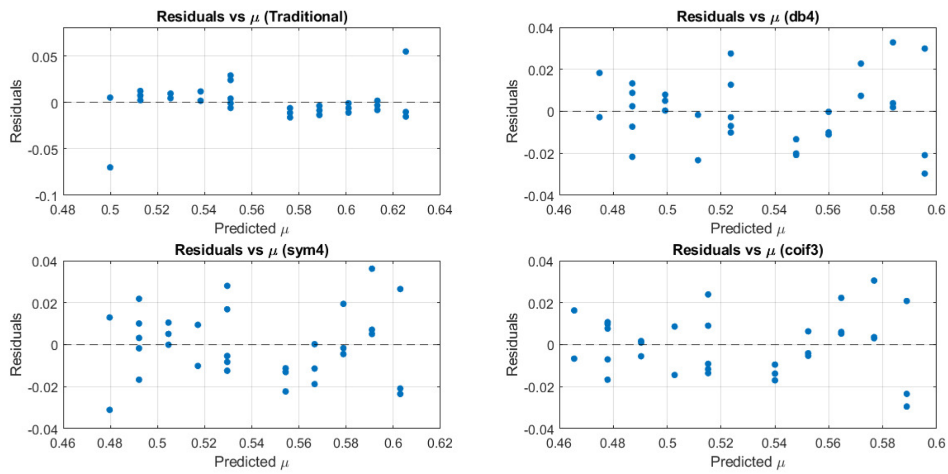

Figure 4. Residual Diagnostics for Traditional and Wavelet-Based Beta Regression Models.

Figure 4 presents residual plots to visualize model adequacy and residual distribution for the four beta regression models. The fit values are plotted on the horizontal axis, and the residual magnitudes on the vertical axis to test dispersion, symmetry, and extreme residuals across models. In the traditional model, the residuals display noticeable dispersion around the fitted line, with a few points deviating substantially, indicating potential model misspecification or unaccounted variability.

Formal likelihood-ratio testing is not explicitly applicable here since the wavelet-enhanced models are not nested extensions of the traditional beta regression model but differ in the estimation of the preprocessed

response. Comparative performance is therefore assessed using information criteria AIC, BIC, MSE, and residual-based diagnostics. On top of that, the graphical checks included normality tests on residuals of deviance and heteroscedasticity checks from residual-fitted relationships to support the visual results.
Against this, wavelet-based models with Db4, Sym4, and Coif4 filters show a greater balance and uniform spread of residuals to zero, with less variance and lower extreme values. Of these, the Coif4 model exhibits the most consistent residual distribution with few outliers and a more tightly clustering pattern, suggesting a better fit of the model and better accommodation of data structure.
This visual evidence complements the diagnostic statistics previously reported, showing that wavelet filtering, especially with Coif4, improves the prediction accuracy of the beta regression model by minimizing the noise and irregularities in the data.

Table 12. Beta Regression Results for Real Data

Model	Traditional	Db4	Sym4	Coif3
Phi	677.52	979.62	931.03	1249
LL	81.454	87.160	86.384	91.017
AIC	-156.91	-168.32	-166.77	-176.03
BIC	-152.51	-163.92	-162.37	-171.64
MSE	0.00035	0.00025	0.00026	0.00020

Table 12 reports the performance metrics for the beta regression models that are applied to the real data set, including the traditional model as well as three models using wavelet filtering: Db4, Sym4, and Coif3. These are Phi, LL, AIC, BIC, and MSE.
Among the three models, Coif3 yields the highest performance in all metrics. It presents, among all, the highest value of Phi (1249), which means that it is better at estimating dispersion as well as the highest log-likelihood (91.017), which means it is a better fit overall. When taken into consideration with other measures to determine the best-fitting model of selecting variables, it also appears that the Coif3 model provides the best compromise between model fit and parsimony, as indicated by the lowest AIC (-176.03) and BIC (-171.64) values compared to the other models. In addition, Coif3 also has the lowest MSE of 0.00020, which signifies a better accuracy of predictions and lower estimation error.
Coif3 performed best in the real data analysis, compared to simulation results. This is due to its greater smoothness and vanishing moments in the wavelet and scaling functions, which help it suppress moderate noise and resolve local irregularities such as outliers. These features enabled Coif3 to better understand the structure of the data on gasoline yield and thus more accurately measure goodness-of-fit.
Outlier sensitivity analyses were conducted to assess the effect of the extreme gas mileage observation in the dataset. in the dataset. The standard beta regressions and the wavelet enhanced models were re-estimated after the outlier was removed and the performance measures of LL, AIC, BIC, MSE were compared to the full dataset. This comparison provides an account of how responsive each model is to local irregularities and determines if the results show improvements centered on the outlier or or whether the response is more robust to noise.
The results presented here validate the overall benefits of wavelet filtering, and particularly the use of the Coif3 wavelet, in providing a more reliable and interpretable beta regression for real data. While, as mentioned above, there is an improvement in fit indices, the increase in predictive accuracy shows the usefulness of the wavelet-based approach in improving regression analyses in complex data situations.
From a statistical perspective, these findings highlight the importance of wavelet family selection when integrating denoising techniques with beta regression models. Practically, the results suggest that Db4 and Sym4 are well-suited for controlled or simulated environments, while Coif3 may be more effective for real-world industrial data characterized by moderate noise and occasional outliers. This distinction provides practical guidance for selecting appropriate wavelet filters in applied beta regression analyses.
From an applied perspective, the proposed wavelet-enhanced beta regression framework offers practical benefits for industrial decision-making and process monitoring. By reducing noise-related variability in bounded quality indicators, the approach enables more reliable estimation of process behavior and improves the detection of subtle changes that may otherwise be masked by random fluctuations. In quality improvement settings, more

stable parameter estimates facilitate informed adjustments to operating conditions, while in process monitoring applications, improved residual behavior enhances the identification of abnormal patterns and potential process deviations. These features make the proposed methodology particularly relevant for industrial environments where data are often noisy, sample sizes are moderate, and timely decisions are required.

4. Limitations

But, in light of positive results there are a number of flaws in this study that must be borne in mind. First, the analysis of real data is based on a very small sample size (32 observations) which may hinder the generalizability of the empirical findings to other application domains. Second, the wavelet denoising procedure uses the universal thresholding rule to ensure methodological consistency and comparability among simulation conditions. A formal sensitivity analysis is not done is a formal sensitivity analysis of alternative thresholding strategies, such as minimax or SURE-based rules, which may provide additional information on the robustness of the denoising process and this is left for further analysis. Third, I used a fixed decomposition level for the discrete wavelet transform and did not explore the potential impact of varying decomposition levels.
On top of that, I deliberately limited the analysis to three wavelet families Db4, Sym4, and Coif3, which were chosen as representative orthogonal wavelets exhibiting complementary localization, symmetry, and smoothness properties. Though other wavelet families such as Haar, Dmey, and Meyer are used widely in practice, their inclusion would greatly expand the scope of this study and could further obscure the interpretation of results. Further analysis of these limitations may be conducted using larger real-world datasets to compare multiple thresholding methods in order to examine adaptive or data-driven levels of decomposition and extend the model to additional wavelet families.

5. Conclusion

Overall, the simulation study had some important conclusions:
1. Wavelet-based denoising greatly improves the fit and stability of beta regressions in noisy data.
2. Both Daubechies and Symlets consistently outperform Coiflets and the conventional model on all measures of performance.
3. Also, larger sample sizes and higher precision parameters provide better model fit and estimate accuracy, while wavelet denoising provides additional benefits.
4. The noise decreases the performance, but wavelet boosted models are far more durable and robust than the alternative method.
Overall, the Real Data Analysis demonstrated the following key conclusions:

1. Wavelet denoising was introduced before the beta regression model. This allowed for significant model fit and predictive accuracy gains over standard beta regression methods.
2. Of all the wavelet filters considered, the Coif3 model consistently performed better than all other models, with the highest values of Phi and log-likelihood, and lowest values of AIC, BIC, and MSE. This indicates improved model fitting and improved prediction accuracy.
3. As a result, residual tests confirmed wavelet-aided models. This indicates that residuals from wavelet-based methods, and particularly the Coif3 model, were more evenly distributed, symmetrically distributed, and tightly clustered around zero. This makes the data structure more accommodating, and there is less hidden variability.
4. 4. All models exhibit a statistically significant relationship between the volume percentage of hydrocarbons and the percentage of gasoline yield, with p-values consistently below 0.001 for both the constant and slope parameters. But the wavelet filtered models, especially Coif3, were more precise estimates with higher Z-statistics and small standard errors.

5. Results indicated that there is an outlier whose yield was 0.680 and affected the dispersion and variability of the data. Yet, the wavelet models dealt with this irregularity better, maintaining the main pattern of data and minimizing the impact of extreme observations.
6. This suggests that wavelet filtering and beta regression would be an appropriate approach to increase robustness and reliability of models in relatively noisy and at times outlier data for industrial processes.
7. The results confirm the practical utility of wavelet-based regression for process monitoring and optimization problems. This is a promising and interesting tool for practitioners who have used continuous and bounded outcome variables.

REFERENCES

1. S. L. P. Ferrari, and F. Cribari-Neto, *Beta regression for modelling rates and proportions*, Journal of Applied Statistics, vol. 31, no. 7, pp. 799–815, 2004.
2. D. Botani, N. Kareem, T. H. Ali, and B. Sedeeq, *Optimizing bandwidth parameter estimation for non-parametric regression using fixed-form threshold with Dmey and Coiflet wavelets*, Hacettepe Journal of Mathematics and Statistics, vol. 54, no. 3, pp. 1094–1106, 2025.
3. H. A. Hayawi, S. M. Azeez, S. O. Babakr, and T. H. Ali, *ARX TIME SERIES MODEL ANALYSIS WITH WAVELETS SHRINKAGE (SIMULATION STUDY)*, Pakistan Journal of Statistics, vol. 41, no. 2, pp. 103–116, 2025.
4. T. H. Ali, A. A. Hamad, S. H. Mahmood, and K. H. Ahmed, *ARIMAX time series analysis for a general budget in the Kurdistan Region of Iraq using wavelet shrinkage*, Communications in Statistics: Case Studies, Data Analysis and Applications, vol. 11, no. 2, pp. 164–188, 2025.
5. A. R. d. S. Sousa, N. L. Garcia, and B. Vidakovic, *Bayesian wavelet shrinkage with beta priors*, Computational Statistics, vol. 36, pp. 1341–1363, 2021.
6. B. Sousa, and J. Zevallos, *Wavelet shrinkage under correlated noise: Bayesian rules and applications*, Statistics and Computing, 2025.
7. S. H. Mahmood, H. A. A. Hayawi, T. H. Ali, B. S. Sedeeq, and S. H. Ali, *Forecasting the CPI of the Kurdistan Region of Iraq Using the Combined ARIMA and GARCH Model With Wavelet Analysis*, Journal of Applied Mathematics, vol. 1, Article ID 2457525, 2025.
8. R. Benhaddou, and Q. Liu, *Wavelet estimation for the nonparametric additive model in random design and long-memory dependent errors*, Journal of Nonparametric Statistics, vol. 36, no. 1, pp. 1–26, 2023.
9. A. W. Omer, and T. H. Ali, *Wavelet Analysis for Outlier Estimation in Multivariate Linear Regression Models*, Passer Journal of Basic and Applied Sciences, vol. 7, no. 1, pp. 478–494, 2025.
10. Y. M. T. Al-Obeady, H. A. A. Hayawi, and M. A. Elkhouli, *Using Wavelets to Identify Linear Dynamic Models*, Iraqi Journal of Statistical Sciences, vol. 22, no. 1, pp. 1–8, 2025.
11. I. I. Elias, and T. H. Ali, *Optimal level and order of the Coiflets wavelet in the VAR time series denoise analysis*, Frontiers in Applied Mathematics and Statistics, vol. 11, 1526540, 2025.
12. H. H. Taha, H. A. A. Hayawi, T. H. Ali, and S. R. Ahmed, *Wavelet Daubechies Enhanced Average Chart Incorporating Classical Shewhart and Bayesian Techniques*, Statistics, Optimization and Information Computing, vol. 14, pp. 2312–2328, 2025.
13. R. Ospina, and S. L. P. Ferrari, *Inflated beta distributions*, Statistical Papers, vol. 51, pp. 111–126, 2010.
14. T. H. Ali, H. A. A. Hayawi, and D. S. I. Botani, *Estimation of the bandwidth parameter in Nadaraya-Watson kernel non-parametric regression based on universal threshold level*, Communications in Statistics - Simulation and Computation, vol. 52, no. 4, pp. 1476–1489, 2023.
15. T. H. Ali, and D. M. Saleh, *Comparison between Wavelet Bayesian and Bayesian Estimators to Remedy Contamination in Linear Regression Model*, PalArch's Journal of Archaeology of Egypt / Egyptology, vol. 18, no. 10, pp. 3388–3409, 2021.
16. J. C. Douma, and J. T. Weedon, *Analyzing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression*, Methods in Ecology and Evolution, vol. 10, no. 9, pp. 1412–1430, 2019.
17. T. H. Ali, H. H. Taha, B. S. Sedeeq, and H. A. A. Hayawi, *An innovative hybrid control chart combining wavelet decomposition and support vector machine for effective outlier detection*, Frontiers in Applied Mathematics and Statistics, vol. 11, 2025.
18. T. H. Ali, *Modification of the adaptive Nadaraya-Watson kernel method for nonparametric regression (simulation study)*, Communications in Statistics - Simulation and Computation, vol. 51, no. 2, pp. 391–403, 2022.
19. R. T. Taha, H. A. A. Hayawi, and M. A. Elkhouli, *Estimation of Delay Time in Linear Dynamic Systems Using Wavelets*, Iraqi Journal of Statistical Sciences, vol. 22, no. 1, pp. 141–150, 2025.
20. T. H. Ali, H. A. A. Hayawi, and H. A. Hamza, *Bayesian time series modelling with wavelet analysis for forecasting monthly inflation*, Iraqi Journal of Statistical Sciences, vol. 22, no. 1, pp. 181–194, 2025.
21. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.
22. S. B. Ameen, and T. H. Ali, *Proposed quality control charts using Haar wavelet coefficients for enhanced production monitoring*, Iraqi Journal of Statistical Sciences, vol. 22, no. 1, pp. 127–140, 2025.
23. D. B. Percival, and A. T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, 2000.
24. M. Misiti, Y. Misiti, G. Oppenheim, and J. M. Poggi, *Wavelet Toolbox User's Guide*, MathWorks, Inc., 2007.
25. M. Al-Hashimi, H. Hayawi, and M. Alawjar, *Ensemble Method for Intervention Analysis to Predict the Water Resources of the Tigris River*, Statistics, Optimization and Information Computing, vol. 14, pp. 144–161, 2025.
26. P. S. Addison, *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, CRC Press, 2017.

27. H. Hayawi, M. Al-Hashimi, and M. Alawjar, *Machine learning methods for modelling and predicting dust storms in Iraq*, Statistics, Optimization and Information Computing, vol. 13, no. 3, pp. 1063–1075, 2025.
28. F. Cribari-Neto, and A. Zeileis, *Beta regression in R*, Journal of Statistical Software, vol. 34, no. 2, pp. 1–24, 2010.
29. M. M. Al-Hashimi, and H. A. Hayawi, *Nonlinear Model for Precipitation Forecasting in Northern Iraq using Machine Learning Algorithms*, International Journal of Mathematics and Computer Science, vol. 19, no. 1, pp. 171–179, 2023.
30. T. A. A.-R. S. Al-Hasso, H. A. A. Hayawi, and T. H. Ali, *Using Linear Wavelets in Analyzing the GARCH Model with the Simulation*, Pakistan Journal of Statistics, vol. 42, no. 1, pp. 27–44, 2026.
31. X. W. XiangJun, and M. M. Al-Hashimi, *The Comparison of Adaptive Neuro-Fuzzy Inference System (ANFIS) with Nonlinear Regression for Estimation and Prediction*, In *2012 International Conference on Information Technology and e-Services*, pp. 1–7, IEEE, 2012.
32. M. M. Y. Al-Hashimi, and X. Wang, *Comparing the Cancer in Ninawa During Three Periods (1980–1990, 1991–2000, 2001–2010) Using Poisson Regression*, Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences, vol. 18, no. 12, p. 1026, 2013.