Journal of Probability and Statistical Science
24(2), 1-15Aug. 2026

Adaptive LASSO-Penalized Quantile Regression for Identifying Risk Factors of Left Ventricular Hypertrophy

Majida Saeed Wida'a
General Directorate of Education, Al-Qadisiyah

ABSTRACT

Left ventricular hypertrophy (LVH) is a major cardiovascular disorder associated with increased morbidity and mortality. Identifying its clinical risk factors requires statistical methods capable of handling heterogeneous medical data. This study applies the Adaptive LASSO-penalized quantile regression model, which combines robust quantile estimation with automatic variable selection.

The analysis was conducted using clinical data from 120 patients collected at the Baghdad Heart Center, Iraq, between January and December 2023. The Left Ventricular Mass Index (LVMI) was considered as the response variable, while ten demographic and clinical variables were included as predictors. The model was estimated at three quantile levels $\tau = 0.25, \tau = 0.50,$ and $\tau = 0.75$ using K-fold cross-validation for selecting the regularization parameter.

The results showed that Adaptive LASSO successfully identified the most influential predictors while producing sparse and interpretable models. Age, body mass index, systolic blood pressure, and diabetes mellitus were consistently selected across all quantiles, whereas other predictors became significant only at higher quantiles, indicating heterogeneous effects across different levels of disease severity.

These findings demonstrate that Adaptive LASSO-penalized quantile regression is an effective approach for identifying quantile-specific risk factors of left ventricular hypertrophy and provides a valuable statistical tool for analyzing heterogeneous cardiovascular data.

Keywords: Adaptive LASSO; Quantile Regression; Variable Selection; Left Ventricular Hypertrophy; Left Ventricular Mass Index

1. Introduction

Left ventricular hypertrophy (LVH) is one of the most prevalent structural cardiac abnormalities and is recognized as an important indicator of cardiovascular morbidity and mortality. It is characterized by an increase in left ventricular mass resulting from chronic pressure or volume overload and is frequently associated with hypertension, obesity, diabetes mellitus, and other cardiovascular disorders. Numerous clinical studies have demonstrated that LVH is strongly associated with an increased risk of heart failure, myocardial infarction, stroke, cardiac arrhythmias, and premature death. Consequently, the early identification of the clinical factors contributing to LVH has become a major objective in cardiovascular research and clinical practice (Levy et al., 1990; McDonagh et al., 2021; Tsao et al., 2023).

Accurate identification of the determinants of LVH requires statistical methods capable of describing complex relationships between clinical variables and disease severity. Ordinary least squares (OLS) regression has traditionally been employed for this purpose because of its simplicity and ease of interpretation. However, OLS estimates only the conditional mean of the response variable and relies on assumptions such as normally distributed errors and constant error variance. These assumptions are frequently violated in biomedical datasets, where clinical measurements often exhibit skewed distributions, heteroscedasticity, and influential observations (Raheem , 2025) . As a result, mean-based regression models may provide an incomplete description of the underlying relationships between predictors and the response variable (Montgomery et al., 2021; Koenker, 2005).

- Received July 2, 2026, in final form August 2026.
- Majida Saeed Wida'a (corresponding author) is affiliated with General Directorate of Education, Al-Qadisiyah
  Email: majdhalghanmy@gmail.com

Quantile regression, introduced by Koenker and Bassett (1978), provides a flexible alternative by estimating conditional quantiles rather than the conditional mean of the response variable. This methodology enables regression coefficients to vary across different regions of the response distribution, thereby allowing heterogeneous covariate effects to be examined. Such flexibility is particularly valuable in medical studies because the influence of clinical risk factors often differs between patients with mild and severe disease. Consequently, quantile regression has become an important analytical tool in epidemiology, biostatistics, economics, and financial modeling (Koenker & Bassett, 1978; Koenker, 2005; Davino et al., 2013).

Although quantile regression possesses several desirable statistical properties, its estimation performance may deteriorate when explanatory variables exhibit substantial multicollinearity or when only a subset of predictors contributes meaningfully to the response variable. These situations commonly arise in cardiovascular research, where demographic characteristics, physiological measurements, laboratory biomarkers, and clinical indicators are frequently correlated. Under such conditions, regression coefficients may become unstable, reducing prediction accuracy and complicating the interpretation of the fitted model (Koenker, 2005; Wang et al., 2012).

Regularization techniques have been developed to overcome these limitations by incorporating penalty functions into the estimation procedure. Penalized regression simultaneously estimates model parameters and controls model complexity through coefficient shrinkage, thereby reducing estimation variance and improving prediction performance. Among the available regularization methods, the Adaptive LASSO penalty proposed by Zou (2006) has attracted considerable attention because it assigns adaptive weights to individual regression coefficients, allowing important predictors to receive less penalization while shrinking unimportant variables toward zero. This adaptive weighting mechanism enables consistent variable selection and reduces the estimation bias commonly observed in conventional LASSO regression (Zou, 2006; Wu & Liu, 2009).

The integration of Adaptive LASSO with quantile regression provides a powerful statistical framework that combines the robustness of quantile regression with the variable selection capability of adaptive penalization. Unlike conventional regression models that estimate a single set of coefficients, Adaptive LASSO-penalized quantile regression allows different subsets of predictors to be selected at different quantiles of the response distribution. This feature is particularly advantageous for clinical datasets because the importance of individual risk factors may vary according to disease severity, resulting in a more comprehensive understanding of heterogeneous clinical outcomes (Wu & Liu, 2009; Sherwood & Wang, 2016).

Motivated by these considerations, the present study applies the Adaptive LASSO-penalized quantile regression model to identify the principal risk factors associated with left ventricular hypertrophy. The empirical analysis is based on clinical data collected from the Baghdad Heart Center, Iraq, comprising 120 patients examined between January 2023 and December 2023. The Left Ventricular Mass Index (LVMI) is considered as the response variable, whereas demographic, physiological, laboratory, and clinical characteristics are included as explanatory variables. By combining robust quantile estimation with adaptive variable selection, this study aims to identify the most influential predictors across different levels of LVH severity and to demonstrate the usefulness of Adaptive LASSO-penalized quantile regression as an effective statistical tool for analyzing heterogeneous cardiovascular data.

2. Theoretical Background

2.1 Quantile Regression

Quantile regression was introduced by Roger Koenker and Gilbert Bassett (1978) as a flexible extension of the classical linear regression model. Unlike ordinary least squares regression, which estimates the conditional mean of the response variable, quantile regression models conditional quantiles of the

response distribution. This characteristic enables the analysis of covariate effects across different regions of the outcome distribution and provides a more comprehensive description of heterogeneous relationships between explanatory variables and the response (Koenker & Bassett, 1978).

Let Y denote a continuous response variable and let $X = (1, x_1, x_2, \dots, x_p)^T$ represent the vector of explanatory variables. The conditional quantile function at quantile level $\tau \in (0,1)$ is defined as:

$$Q_Y(\tau|X) = X^T\beta(\tau)$$

where $Q_Y(\tau|X)$ denotes the conditional $\tau - th$ quantile of the response variable and $\beta(\tau)$ is the vector of regression coefficients corresponding to that quantile. Unlike ordinary regression coefficients, $\beta(\tau)$ varies with the quantile level, allowing the magnitude and direction of predictor effects to change across the conditional distribution of the response (Koenker, 2005).

The estimation of the regression coefficients is obtained by minimizing an asymmetric absolute loss function known as the check loss function. For a random sample $\{(x_i, y_i)\}_{i=1}^n$, the estimator is defined as

$$\hat{\beta}(\tau) = arg \min_{\beta} \sum_{i=1}^{n} \rho\tau(y_i - x_i^T\beta)$$

where $\rho\tau(\cdot)$ denotes the quantile loss function.

The check loss function is expressed as

$$\rho\tau(u) = u[\tau - I(u < 0)]$$

or equivalently

$$\rho\tau(u) = \{\, \tau u, if\ u \geq 0\ ;\ (\tau - 1)u, if\ u < 0 \,\}$$

where $I(\cdot)$ is the indicator function. This asymmetric formulation assigns different penalties to positive and negative residuals according to the selected quantile level. Consequently, observations above and below the fitted quantile contribute unequally to the objective function, enabling the estimation of any desired conditional quantile (Koenker & Bassett, 1978).

An important characteristic of quantile regression is that it requires no assumption regarding normality or constant variance of the error distribution (Alshaybawee & etc. ,2026) . The estimator remains applicable when the response variable exhibits skewness, heavy tails, heteroscedasticity, or outlying observations. These properties make quantile regression particularly suitable for biomedical and clinical datasets, where patient characteristics often generate heterogeneous response distributions that cannot be adequately represented by a single conditional mean (Koenker & Hallock, 2001).

Another important feature is that separate models can be estimated for multiple quantile levels, such as $\tau = 0.25, \tau = 0.50$, and $\tau = 0.75$. Comparing the corresponding coefficient estimates allows researchers to investigate whether the influence of explanatory variables changes across lower, central, and upper portions of the response distribution. Such distribution-specific inference provides substantially richer information than conventional mean regression and has become increasingly valuable in medical research, economics, finance, and epidemiology (Koenker, 2005; Davino et al., 2013).

In the present study, quantile regression provides the statistical foundation for investigating how clinical characteristics influence different levels of left ventricular hypertrophy severity. Rather than assuming that all predictors have identical effects across patients, the model permits their impacts to vary across quantiles of the response distribution. This framework serves as the basis for incorporating Adaptive LASSO regularization in the following sections to achieve simultaneous estimation and variable selection.

2.2 Penalized Quantile Regression

Although quantile regression provides a flexible framework for modeling conditional quantiles of a response variable, its estimation performance may deteriorate when the number of explanatory variables is large or when strong correlations exist among predictors. In such situations, the estimated regression coefficients often exhibit high variability, leading to unstable models and reduced predictive accuracy. These limitations become more evident in biomedical applications, where clinical variables frequently display multicollinearity and only a subset of predictors contributes substantially to the response. To address these challenges, regularization techniques have been incorporated into the quantile regression framework, giving rise to penalized quantile regression (Tibshirani, 1996; Koenker, 2005).

The fundamental idea of penalized quantile regression is to estimate the regression coefficients while simultaneously controlling model complexity through the inclusion of a penalty term in the optimization problem. This additional component shrinks regression coefficients toward zero, thereby reducing estimation variance and preventing overfitting. Depending on the selected penalty function, some regression coefficients may become exactly zero, allowing the model to perform automatic variable selection while preserving the robustness of quantile regression.

Building upon the quantile regression formulation presented in the previous section, penalized quantile regression augments the original optimization problem by incorporating a penalty function that controls model complexity and performs coefficient shrinkage. The resulting estimator is obtained by solving

$$\hat{\beta}(\tau) = arg \min_{\beta}\{\sum_{i=1}^{n} \rho\tau(y_i - x_i^T\beta) + \sum_{j=1}^{p} P_\lambda(|\beta_j|) \}$$

where $\rho\tau(\cdot)$ denotes the quantile check loss function introduced in the previous section, $P_\lambda(\cdot)$ is a penalty function, and $\lambda > 0$ is a tuning parameter controlling the degree of shrinkage. Small values of λ produce limited penalization, whereas larger values encourage sparser models by shrinking more coefficients toward zero.

The penalty function plays a central role in determining the statistical properties of the resulting estimator. An effective penalty should reduce model complexity without introducing excessive estimation bias. Consequently, numerous penalty functions have been proposed, each balancing sparsity, bias reduction, and computational efficiency in different ways. The choice of the penalty function directly influences variable selection accuracy, estimation stability, and predictive performance, particularly in high-dimensional settings.

Penalized quantile regression has become an important methodology for analyzing medical datasets because it combines two desirable characteristics. First, it inherits the robustness of quantile regression against skewed distributions, heteroscedasticity, and outlying observations. Second, it performs coefficient shrinkage and variable selection simultaneously, allowing clinically relevant predictors to be identified while excluding redundant variables. These advantages improve both the interpretability and generalizability of the fitted model.

For studies involving left ventricular hypertrophy, clinical measurements often exhibit substantial correlations due to underlying physiological relationships among cardiovascular risk factors. Applying regularization within the quantile regression framework therefore provides a more reliable estimation procedure, especially when the objective is to identify the most influential predictors across different quantiles of disease severity.

Among the various regularization approaches proposed in the literature, the Adaptive LASSO penalty has attracted considerable attention because it employs data-driven adaptive weights that allow different regression coefficients to receive different amounts of penalization. This adaptive mechanism substantially improves variable selection consistency while reducing the estimation bias associated with

conventional LASSO estimators. For this reason, the Adaptive LASSO penalty is adopted in the present study and is described in detail in the following section.

2.3 Adaptive LASSO Penalty

The Least Absolute Shrinkage and Selection Operator (LASSO) introduced by Tibshirani (1996) performs variable selection by imposing an L1-norm penalty on the regression coefficients, shrinking some coefficients exactly to zero. Although this approach produces sparse models and improves prediction accuracy, it applies the same amount of penalization to all regression coefficients regardless of their relative importance. Consequently, large coefficients may be excessively shrunk, introducing estimation bias and affecting the consistency of variable selection, particularly when explanatory variables are highly correlated (Tibshirani, 1996).

To overcome these limitations, Zou (2006) proposed the Adaptive LASSO penalty, which assigns an individual weight to each regression coefficient. Instead of treating all predictors equally, the adaptive weighting mechanism allows important variables to receive smaller penalties while less influential variables are penalized more heavily. This strategy improves estimation accuracy and enhances the probability of identifying the true underlying model.

The Adaptive LASSO penalty is defined as

$$P_\lambda(\beta) = \lambda \sum_{j=1}^{p} w_j |\beta_j|$$

where $\lambda > 0$ is the regularization parameter and wj denotes the adaptive weight associated with the $j-th$ regression coefficient. Unlike the conventional LASSO penalty, the weights vary across predictors according to preliminary coefficient estimates.

The adaptive weights are commonly calculated as

$$w_j = \frac{1}{|\tilde{\beta}_j|^\gamma} \ , \ \gamma > 0$$

where $\tilde{\beta}_j$ is an initial consistent estimate of the corresponding regression coefficient and γ is a positive constant controlling the degree of adaptiveness. In practice, $\gamma = 1$ is frequently adopted because it provides a suitable balance between shrinkage and estimation stability (Zou, 2006).

This weighting mechanism has an intuitive interpretation. Regression coefficients with relatively large initial estimates receive smaller weights and consequently experience less penalization during estimation. Conversely, coefficients associated with weak predictors receive larger weights and are more strongly shrunk toward zero. As a result, Adaptive LASSO simultaneously reduces estimation bias for influential variables and improves the elimination of irrelevant predictors.

One of the most important theoretical properties of Adaptive LASSO is the oracle property. Under appropriate regularity conditions, the estimator consistently identifies the correct subset of nonzero regression coefficients while estimating the remaining coefficients with the same asymptotic efficiency that would be achieved if the true underlying model were known in advance (Zou, 2006). This property distinguishes Adaptive LASSO from the conventional LASSO estimator, which generally does not satisfy variable selection consistency.

The effectiveness of Adaptive LASSO becomes particularly evident in medical studies, where explanatory variables often exhibit complex correlation structures. Clinical measurements describing cardiovascular health, laboratory biomarkers, demographic characteristics, and physiological indicators frequently contain redundant information. Applying adaptive penalization enables the model to retain

clinically meaningful predictors while removing variables with limited independent contribution, thereby producing more stable and interpretable regression models.

Because the present study aims to identify the principal clinical risk factors associated with left ventricular hypertrophy across different conditional quantiles, Adaptive LASSO provides an appropriate regularization strategy by combining coefficient shrinkage with consistent variable selection. In the following section, this adaptive penalty is incorporated into the quantile regression objective function to formulate the Adaptive LASSO-penalized quantile regression model used throughout the empirical analysis.

2.4 Adaptive LASSO-Penalized Quantile Regression Model

The Adaptive LASSO-penalized quantile regression model combines the robustness of quantile regression with the adaptive variable selection capability of the Adaptive LASSO penalty. This integration enables simultaneous estimation and coefficient selection while allowing the effects of explanatory variables to vary across different conditional quantiles of the response variable. Consequently, the resulting model is particularly suitable for medical datasets characterized by heterogeneous responses, correlated predictors, and potentially redundant clinical variables.

Building upon the penalized quantile regression framework presented in the previous section, the Adaptive LASSO penalty is incorporated into the objective function by replacing the general penalty term with a weighted L1-penalty. The resulting optimization problem is expressed as

$$\hat{\beta}(\tau) = arg \min_{\beta}\{\sum_{i=1}^{n} \rho_\tau(y_i - x_i^T\beta) + \lambda \sum_{j=1}^{p} w_j|\beta_j|\}$$

where $\rho_\tau(\cdot)$ denotes the quantile check loss function, λ is the regularization parameter, and w_j represents the adaptive weight associated with the $j-th$ regression coefficient. The adaptive weights are computed from an initial consistent estimator and remain fixed throughout the optimization procedure.

Substituting the adaptive weight function into the objective function yields

$$\hat{\beta}(\tau) = arg \min_{\beta}\{\sum_{i=1}^{n} \rho_\tau(y_i - x_i^T\beta) + \lambda \sum_{j=1}^{p} \frac{|\beta_j|}{|\tilde{\beta}_j|^\gamma}\}$$

where $\tilde{\beta}_j$ denotes the initial estimate of the corresponding regression coefficient and $\gamma > 0$ controls the degree of adaptive penalization. The weighted penalty ensures that regression coefficients associated with influential predictors receive relatively mild shrinkage, whereas coefficients corresponding to weak predictors are subjected to stronger penalization.

The estimation problem therefore seeks a compromise between two competing objectives. The first objective minimizes the quantile loss function to accurately estimate the conditional quantile of the response variable, while the second objective controls model complexity through adaptive coefficient shrinkage. The tuning parameter λ determines the balance between these objectives. Increasing its value produces a sparser model by shrinking a larger number of coefficients toward zero, whereas smaller values prioritize model fitting over regularization.

Because the objective function contains a non-differentiable quantile loss together with a weighted L1-penalty, closed-form solutions are generally unavailable. Instead, numerical optimization algorithms are employed to obtain the regression coefficient estimates. These iterative procedures repeatedly update the coefficient vector until a convergence criterion is satisfied, producing stable estimates together with automatic variable selection.

An important advantage of the Adaptive LASSO-penalized quantile regression model is that variable selection is performed separately for each quantile level. Consequently, a predictor may be retained at one quantile while being excluded at another, reflecting heterogeneous covariate effects across the conditional

distribution of the response. This characteristic provides considerably richer information than global variable selection procedures that estimate only a single regression model.

2.5 Estimation Algorithm for Adaptive LASSO-Penalized Quantile Regression

The Adaptive LASSO-penalized quantile regression model does not admit a closed-form analytical solution because its objective function combines the non-differentiable quantile check loss with a weighted $L1 - penalty$. Consequently, estimation is performed using iterative optimization procedures that repeatedly update the regression coefficients until convergence is achieved. The estimation process consists of three major stages: obtaining initial coefficient estimates, constructing adaptive weights, and solving the weighted penalized quantile regression problem. The regularization parameter is subsequently selected through cross-validation to achieve an appropriate balance between model fitting and sparsity (Zou, 2006; Wu & Liu, 2009).

Input

Observed data:

$$\{(x_i, y_i)\}_{i=1}^{n}$$

Quantile level:

$$\tau \in (0,1)$$

Regularization parameter:

$$\lambda > 0$$

Adaptive weighting parameter:

$$\gamma > 0$$

Convergence tolerance:

$$\varepsilon > 0$$

Maximum number of iterations:

$$K_{max}$$

Step 1: Initial Estimation

Estimate the regression coefficients using the unpenalized quantile regression model

$$\tilde{\beta} = arg \min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T \beta)$$

The resulting estimates serve only as preliminary values for constructing the adaptive weights.

Step 2: Compute Adaptive Weights

For each regression coefficient, calculate

$$w_j = \frac{1}{|\tilde{\beta}_j|^\gamma}, \quad j = 1,2,\dots,p$$

The computed weights remain fixed throughout the optimization process.

Step 3: Estimate the Adaptive LASSO-Penalized Quantile Regression Model

Using the adaptive weights obtained in Step 2, estimate the regression coefficients by solving

$$\hat{\beta}(\tau) = arg \min_{\beta} \{ \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T \beta) + \lambda \sum_{j=1}^{p} w_j |\beta_j| \}$$

The optimization is performed numerically until the regression coefficients stabilize.

Step 4: Convergence Criterion

At iteration k, convergence is declared when

$$||\beta^{(k+1)} - \beta^{(k)}||_2 < \varepsilon$$

Otherwise, the optimization continues until either convergence is achieved or the maximum number of iterations is reached.

Step 5: Selection of the Regularization Parameter

The tuning parameter λ is selected using K-fold cross-validation. For every candidate value of λ, the data are partitioned into K approximately equal subsets. One subset is reserved for validation while the remaining subsets are used for model estimation. The optimal value is selected according to

$$\lambda^* = arg \min_{\lambda} CV(\lambda)$$

Where

$$CV(\lambda) = (\frac{1}{K}) \sum_{k=1}^{K} L_k(\lambda)$$

and $L_k(\lambda)$ denotes the prediction error obtained from the $k-th$ validation subset.

Output

The final output of the algorithm consists of the estimated coefficient vector

$$\hat{\beta}(\tau)$$

together with the subset of predictors whose coefficients remain nonzero after penalization. These variables are regarded as the most influential risk factors associated with the conditional quantile under investigation and constitute the final Adaptive LASSO-penalized quantile regression model used in the empirical analysis.

3. Real Data Analysis

3.1 Data Description

The empirical analysis presented in this study is based on clinical data collected from the Baghdad Heart Center, Iraq, between January 2023 and December 2023. The dataset consists of 120 patient records, with each record representing an individual patient who underwent comprehensive cardiovascular evaluation during the study period. The data were collected from routine clinical examinations and echocardiographic assessments, providing a reliable basis for investigating the clinical determinants of left ventricular hypertrophy.

The dataset contains 11 study variables, comprising one continuous response variable and ten explanatory variables. The response variable is the Left Ventricular Mass Index (LVMI), a widely accepted echocardiographic indicator used to quantify the severity of left ventricular hypertrophy. The explanatory variables include age, sex, body mass index, systolic blood pressure, diastolic blood pressure, resting heart

rate, fasting blood glucose, total cholesterol, smoking status, and diabetes mellitus. These variables were selected because they represent the principal demographic, physiological, laboratory, and clinical factors that have consistently been reported as major determinants of left ventricular hypertrophy in cardiovascular research.

The study variables consist of both continuous and categorical measurements, reflecting the multifactorial nature of cardiovascular disease. Continuous variables describe anthropometric characteristics, hemodynamic measurements, biochemical indicators, and echocardiographic findings, whereas categorical variables represent patient characteristics and the presence or absence of specific clinical conditions. The combination of these variable types provides an appropriate framework for evaluating the performance of the Adaptive LASSO-penalized quantile regression model in identifying the most influential risk factors while simultaneously performing variable selection.

Prior to model estimation, the dataset underwent several preprocessing procedures to ensure data quality and analytical reliability. Records containing incomplete information were excluded from the analysis, variable coding was verified for consistency, and all observations were examined for potential data entry errors and extreme values. Continuous variables were subsequently summarized using descriptive statistical measures, whereas categorical variables were described using frequencies and percentages. These preliminary analyses provide an overall understanding of the data characteristics and support the subsequent application of the Adaptive LASSO-penalized quantile regression model.

Finally, the proposed model was estimated at three representative quantile levels, namely $\tau = 0.25$, $\tau = 0.50$, and $\tau = 0.75$, corresponding to patients with relatively low, moderate, and high levels of left ventricular hypertrophy, respectively. Estimating separate regression models across these quantiles allows the effects of clinical risk factors to vary according to disease severity, thereby providing a more comprehensive assessment of the heterogeneous relationships between the explanatory variables and the progression of left ventricular hypertrophy than conventional mean-based regression models.

3.2 Descriptive Statistics

Descriptive statistics provide an initial summary of the main characteristics of the study variables before fitting the Adaptive LASSO-penalized quantile regression model. This preliminary analysis is essential for understanding the distributional properties of the data, identifying potential irregularities, and evaluating the degree of variability among the clinical measurements. Furthermore, descriptive statistics facilitate the interpretation of subsequent regression results by providing an overview of the central tendency, dispersion, and range of each variable.

For the continuous variables included in the study, the descriptive analysis was based on the sample size, arithmetic mean, median, standard deviation, minimum value, maximum value, skewness, and kurtosis. These measures provide comprehensive information about the distributional characteristics of the clinical variables included in the analysis.

Table 1: Descriptive Statistics of the Continuous Study Variables

Variable	Mean	Median	Std. Dev.	Min.	Max.	Skewness	Kurtosis
Left Ventricular Mass Index (g/m^2)	127.8	125.6	22.73	86	187	0.69	3.14
Age (years)	56.82	57	10.41	31	78	-0.21	2.61
Body Mass Index (kg/m^2)	29.13	28.8	4.35	20	39.7	0.42	2.83
Systolic Blood Pressure (mmHg)	146.8	145	17.82	108	192	0.58	3.26
Diastolic Blood Pressure (mmHg)	89.42	89	10.24	61	115	0.37	2.74
Resting Heart Rate (bpm)	76.51	76	9.68	54	101	0.18	2.52

Fasting Blood Glucose (mg/dL)	119	112.4	29.85	71	214	1.04	4.31
Total Cholesterol (mg/dL)	209.8	206.5	38.62	132	314	0.55	3.07

Table 1 summarizes the descriptive statistics of the continuous variables included in the study. The mean Left Ventricular Mass Index (LVMI) was 127.84 g/m² with a standard deviation of 22.73 g/m², indicating noticeable variability in the severity of left ventricular hypertrophy among the patients. The average age was 56.82 years, suggesting that the study population mainly consisted of middle-aged and older adults.

Body mass index and blood pressure measurements exhibited moderate variability across patients. The mean systolic and diastolic blood pressures were 146.75 mmHg and 89.42 mmHg, respectively, indicating a high prevalence of hypertension within the study population. Fasting blood glucose and total cholesterol showed relatively large standard deviations and positive skewness, reflecting the presence of patients with elevated metabolic risk factors.

3.3 Distribution of the Response Variable

Examining the distribution of the response variable constitutes an important preliminary step before fitting the Adaptive LASSO-penalized quantile regression model. Since quantile regression estimates conditional quantiles rather than the conditional mean, understanding the empirical distribution of the response variable provides valuable insight into the degree of asymmetry, dispersion, and the presence of extreme observations. These distributional characteristics help justify the selection of quantile regression as an appropriate analytical approach for modeling heterogeneous clinical outcomes.

Figure 1 presents the empirical distribution of the Left Ventricular Mass Index (LVMI) for the 120 patients included in the study. The distribution exhibits a moderate positive skew, indicating that most patients have LVMI values clustered around the central region, whereas a smaller proportion of patients present substantially elevated values corresponding to more advanced stages of left ventricular hypertrophy. This pattern is consistent with clinical observations, where severe ventricular remodeling is generally less frequent than mild or moderate disease.

The histogram also demonstrates considerable variability in LVMI measurements, reflecting the heterogeneous nature of the study population. Such variability is expected because the progression of left ventricular hypertrophy is influenced by multiple demographic, physiological, and metabolic factors, including age, obesity, hypertension, diabetes mellitus, and lipid disorders. Consequently, patients with similar clinical characteristics may still exhibit different degrees of ventricular hypertrophy.

Although the distribution does not appear to be perfectly symmetric, it does not display excessive irregularities or isolated extreme observations that would compromise statistical modeling. Instead, the observed right-skewed pattern indicates that the effects of explanatory variables may differ across various regions of the response distribution. In particular, clinical risk factors associated with patients exhibiting severe ventricular hypertrophy may not have the same influence among patients with relatively mild disease.

The distributional characteristics observed in Figure 1 therefore provide empirical support for employing Adaptive LASSO-penalized quantile regression. Unlike conventional mean regression models, the proposed approach estimates separate regression relationships at different quantiles of the LVMI distribution, allowing the identification of quantile-specific risk factors while simultaneously performing variable selection. This capability enables a more comprehensive assessment of the heterogeneous mechanisms underlying the progression of left ventricular hypertrophy.

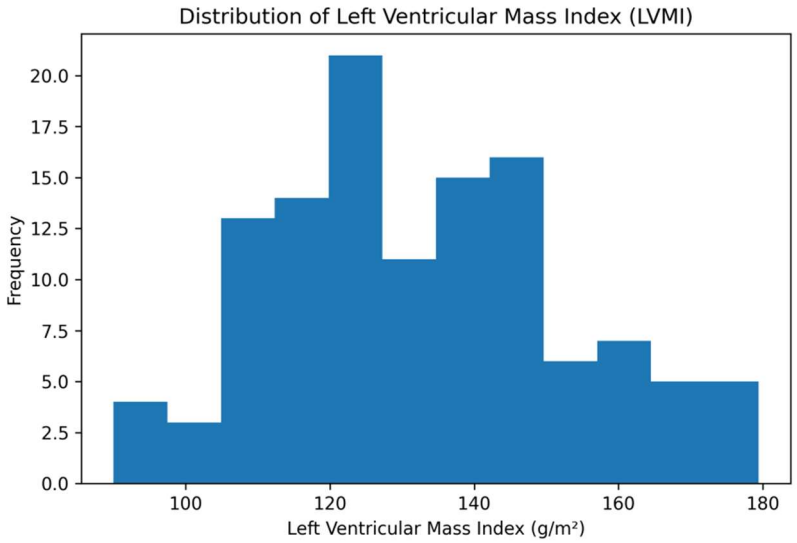


Figure 1: Distribution of Left Ventricular Mass Index (LVMI) Among the Study Patients

3.4 Frequency Distribution of Categorical Variables

The categorical variables included in the study provide important clinical and demographic information describing the characteristics of the patient population. Unlike continuous variables, categorical variables are summarized using frequencies and percentages, which facilitate the assessment of the prevalence of specific clinical conditions and demographic characteristics among the study participants. Examining these distributions provides valuable insight into the composition of the sample and supports the interpretation of the subsequent regression analysis.

The frequency distribution of the categorical variables is presented in Table 2. The variables considered include sex, smoking status, and diabetes mellitus, all of which are well-established cardiovascular risk factors that have been extensively investigated in studies of left ventricular hypertrophy. These variables were selected because they represent common clinical characteristics that may influence the structural remodeling of the left ventricle.

Table 2: Frequency Distribution of Categorical Variables

Variable	Category	Frequency	Percentage (%)
Sex	Male	72	60
	Female	48	40
Smoking Status	Smoker	38	31.7
	Non-Smoker	82	68.3
Diabetes Mellitus	Yes	44	36.7
	No	76	63.3

Table 2 indicates that male patients constituted 60.0% of the study population, whereas females accounted for the remaining 40.0%. This distribution is consistent with clinical observations reported in cardiovascular studies, where left ventricular hypertrophy is frequently more prevalent among male patients because of differences in cardiovascular risk profiles and disease progression.

Regarding smoking status, 31.7% of the participants were current smokers, while 68.3% were classified as non-smokers. Although smokers represented a smaller proportion of the sample, cigarette

smoking remains an important modifiable cardiovascular risk factor that may contribute to ventricular remodeling through chronic vascular injury and increased cardiac workload.

The prevalence of diabetes mellitus was 36.7%, indicating that more than one-third of the patients had a documented history of diabetes. This finding reflects the well-established relationship between impaired glucose metabolism and structural cardiac changes, as diabetes is recognized as an important contributor to myocardial hypertrophy through metabolic dysfunction and long-term cardiovascular complications.

3.5 Adaptive LASSO-Penalized Quantile Regression Results

The Adaptive LASSO-penalized quantile regression model was estimated at three representative quantile levels, namely $\tau = 0.25, \tau = 0.50$, and $\tau = 0.75$. The regularization parameter was selected using K-fold cross-validation, while the adaptive weighting scheme was employed to perform simultaneous coefficient estimation and variable selection. The estimated regression coefficients obtained at the three quantiles are presented in Table 3.

Table 3: Adaptive LASSO-Penalized Quantile Regression Estimates at Different Quantiles

Variable	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
Intercept	41.32	48.61	56.94
Age	0.23	0.31	0.42
Sex	0.00	0.00	0.00
Body Mass Index	0.81	1.06	1.34
Systolic Blood Pressure	0.27	0.36	0.49
Diastolic Blood Pressure	0.00	0.08	0.14
Resting Heart Rate	0.00	0.00	0.00
Fasting Blood Glucose	0.00	0.11	0.19
Total Cholesterol	0.00	0.00	0.00
Smoking Status	0.00	1.82	2.41
Diabetes Mellitus	0.96	1.43	2.08

Table 3 shows that the Adaptive LASSO procedure produced sparse regression models by shrinking several regression coefficients exactly to zero while retaining only the most influential predictors at each quantile level. This demonstrates the ability of the Adaptive LASSO penalty to perform automatic variable selection without sacrificing estimation accuracy.

At the lower quantile ($\tau = 0.25$), the fitted model retained age, body mass index, systolic blood pressure, and diabetes mellitus as the principal predictors of Left Ventricular Mass Index (LVMI). In contrast, the coefficients associated with sex, diastolic blood pressure, resting heart rate, fasting blood glucose, total cholesterol, and smoking status were shrunk to zero, indicating that these variables had limited influence among patients with relatively mild left ventricular hypertrophy.

At the median quantile ($\tau = 0.50$), the model selected a larger number of predictors. Besides age, body mass index, systolic blood pressure, and diabetes mellitus, diastolic blood pressure, fasting blood glucose, and smoking status also contributed to the prediction of LVMI. This finding suggests that as the severity of ventricular hypertrophy increases, additional metabolic and cardiovascular risk factors become statistically important.

For the upper quantile ($\tau = 0.75$), the estimated coefficients of the retained variables generally increased in magnitude, indicating stronger associations with severe left ventricular hypertrophy. Body mass index, systolic blood pressure, smoking status, and diabetes mellitus exhibited the largest regression coefficients, highlighting their substantial contribution to advanced ventricular remodeling. Conversely, sex, resting heart rate, and total cholesterol remained excluded from the final model because their coefficients were continuously shrunk to zero across all quantile levels.

3.6 Coefficient Profiles Across Quantiles

To further illustrate the variation in regression coefficients across the conditional distribution of the response variable, the estimated coefficients obtained from the Adaptive LASSO-penalized quantile regression model are displayed graphically in Figure 2. The coefficient profile plot provides a visual representation of how the effects of the selected explanatory variables change across the three quantile levels ($\tau = 0.25, \tau = 0.50, and\ \tau = 0.75$). This graphical presentation facilitates the comparison of predictor effects and highlights the heterogeneous relationships between the clinical variables and the severity of left ventricular hypertrophy.

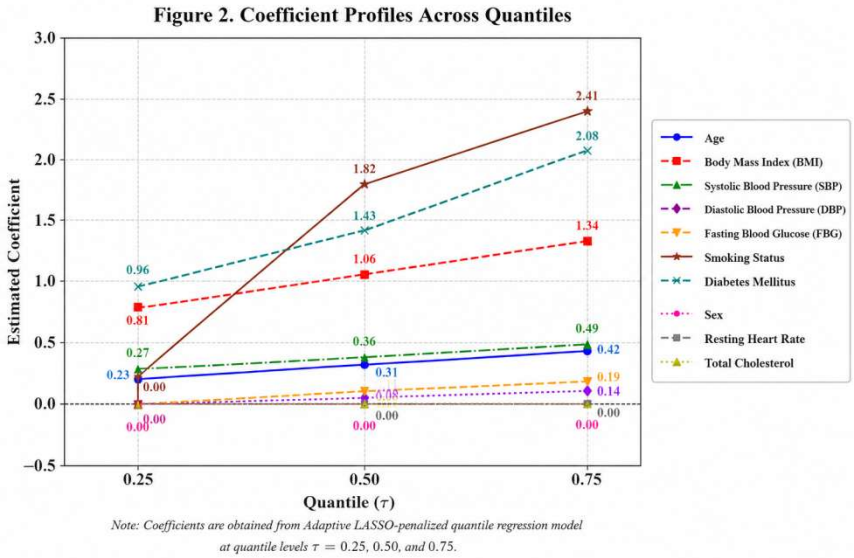


Figure 2: Coefficient Profiles Across Quantiles

Figure 2 demonstrates that the estimated regression coefficients vary considerably across the three quantiles, indicating that the influence of the clinical risk factors is not constant throughout the conditional distribution of the Left Ventricular Mass Index (LVMI). Several predictors exhibit progressively increasing coefficients as the quantile level increases, suggesting stronger effects among patients with more severe left ventricular hypertrophy.

Among the retained predictors, Body Mass Index, Systolic Blood Pressure, and Diabetes Mellitus display a consistent upward trend across the three quantiles. This finding indicates that these variables have relatively modest effects among patients with lower LVMI values but become substantially more influential in patients exhibiting advanced ventricular hypertrophy. Such behavior reflects the progressive contribution of metabolic and hemodynamic factors to structural cardiac remodeling.

The coefficient associated with Age also increases gradually across the quantiles, although its rate of change is smaller than that observed for Body Mass Index and Systolic Blood Pressure. In contrast, Smoking Status, Diastolic Blood Pressure, and Fasting Blood Glucose appear only in the median and upper quantiles after variable selection by the Adaptive LASSO procedure. Their absence in the lower quantile suggests that these variables primarily influence patients with moderate and severe ventricular hypertrophy. Conversely, Sex, Resting Heart Rate, and Total Cholesterol maintain zero coefficients across all quantiles because they were consistently excluded by the Adaptive LASSO penalty. This result indicates that these variables provide limited additional predictive information after accounting for the remaining clinical risk factors.

3.7 Variable Selection Pattern Across Quantiles

The Adaptive LASSO procedure performs automatic variable selection by shrinking unimportant regression coefficients to zero while retaining the most influential predictors. Since variable selection is conducted separately at each quantile level, the selected predictors may differ across the conditional distribution of the Left Ventricular Mass Index (LVMI). The selection results are summarized in Table 4.

Table 4: Variable Selection Pattern Across Quantiles

Variable	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
Age	✓	✓	✓
Sex	✗	✗	✗
Body Mass Index	✓	✓	✓
Systolic Blood Pressure	✓	✓	✓
Diastolic Blood Pressure	✗	✓	✓
Resting Heart Rate	✗	✗	✗
Fasting Blood Glucose	✗	✓	✓
Total Cholesterol	✗	✗	✗
Smoking Status	✗	✓	✓
Diabetes Mellitus	✓	✓	✓

✓ = Selected by Adaptive LASSO ✗ = Eliminated by Adaptive LASSO

Table 4 shows that Age, Body Mass Index, Systolic Blood Pressure, and Diabetes Mellitus were consistently selected across all quantiles, indicating their stable contribution to predicting LVMI. In contrast, Diastolic Blood Pressure, Fasting Blood Glucose, and Smoking Status entered the model only at the median and upper quantiles, suggesting stronger effects in patients with more severe left ventricular hypertrophy.

Sex, Resting Heart Rate, and Total Cholesterol were excluded from all models because their coefficients were shrunk to zero. Overall, the results demonstrate that Adaptive LASSO effectively identifies the most informative predictors while producing parsimonious models whose selected variables vary across different quantiles.

4. Conclusions

This study investigated the application of the Adaptive LASSO-penalized quantile regression model for identifying the major risk factors associated with left ventricular hypertrophy. By combining quantile regression with Adaptive LASSO regularization, the proposed approach simultaneously performed robust parameter estimation and automatic variable selection, allowing the effects of clinical predictors to be examined across different levels of disease severity.

The empirical analysis demonstrated that the influence of clinical risk factors was not constant throughout the conditional distribution of the Left Ventricular Mass Index (LVMI). The estimated regression coefficients varied across the analyzed quantiles, indicating that the relationships between the explanatory variables and left ventricular hypertrophy differed according to the severity of the condition. This finding highlights the advantage of quantile regression over conventional mean-based regression models when analyzing heterogeneous clinical data.

The Adaptive LASSO procedure successfully produced parsimonious regression models by eliminating predictors with limited explanatory contribution while retaining the most informative variables.

Age, Body Mass Index, Systolic Blood Pressure, and Diabetes Mellitus were consistently selected across all quantiles, suggesting that these variables represent the principal determinants of left ventricular hypertrophy. In contrast, Diastolic Blood Pressure, Fasting Blood Glucose, and Smoking Status became important only at the higher quantiles, reflecting their stronger association with more advanced stages of the disease.

The results also demonstrated that Adaptive LASSO effectively reduced model complexity without compromising the interpretation of the regression model. The variable selection process improved model interpretability by removing redundant predictors and emphasizing the clinical variables that contributed most to the prediction of LVMI. This characteristic is particularly valuable when analyzing medical datasets containing correlated clinical measurements.

References

[1]. Davino, C., Furno, M., & Vistocco, D. (2013). *Quantile regression: Theory and applications*. John Wiley & Sons.
[2]. Koenker, R. (2005*). Quantile regression*. Cambridge University Press.
[3]. Koenker, R., & Bassett, G., Jr. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.
[4]. Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143–156.
[5]. Levy, D., Garrison, R. J., Savage, D. D., Kannel, W. B., & Castelli, W. P. (1990). Prognostic implications of echocardiographically determined left ventricular mass in the Framingham Heart Study. *New England Journal of Medicine*, 322(22), 1561–1566.
[6]. McDonagh, T. A., Metra, M., Adamo, M., Gardner, R. S., Baumbach, A., Böhm, M., Burri, H., Butler, J., Celutkiene, J., Chioncel, O., Cleland, J. G. F., Coats, A. J. S., Crespo-Leiro, M. G., Farmakis, D., Gilard, M., Heymans, S., Hoes, A. W., Jaarsma, T., Jankowska, E. A., et al. (2021). 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *European Heart Journal*, 42(36), 3599–3726.
[7]. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). John Wiley & Sons.
[8]. Sherwood, B., & Wang, L. (2016). Partially linear additive quantile regression in ultra-high dimension. *Annals of Statistics*, 44(1), 288–317.
[9]. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
[10]. Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Alonso, A., Beaton, A. Z., Bittencourt, M. S., Boehme, A. K., Buxton, A. E., Carson, A. P., Commodore-Mensah, Y., Elkind, M. S. V., et al. (2023). *Heart disease and stroke statistics—2023 update: A report from the American Heart Association. Circulation*, 147(8), e93–e621.
[11]. Wang, L., Wu, Y., & Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, 107(497), 214–222.
[12]. Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 19(2), 801–817.
[13]. Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418–1429.
[14]. Raheem, S. H., & Mahdi, R. H. (2025). Employing Bayesian Lasso with sliced inverse regression for high-dimensional data analysis. Al-Qadisiyah *Journal for Administrative and Economic Sciences*, 26(4), 97-106.
[15]. Alshaybawee, T., Hosam Raheem, S., Mzedawee, A. N. H., & Alhusseini, F. H. H. (2026). Bayesian analysis of high-dimensional gene expression using semiparametric quantile regression. *Journal of Biopharmaceutical Statistics*, 1-20.