ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Non-Linear Data Modeling using Regression Techniques

Zhwan Rashid Aziz*, **Samira Muhamad Salh**

Department of Statistical and Information, College of Administration and Economic, Sulaimani University, Sulaymaniyah, Iraq

ABSTRACT

The purpose of this study is to use multiple linear and polynomial regression techniques to examine the relation between blood pressure (BP) and certain biochemical markers. For this study, 180 observations from a cross-sectional dataset were examined. Blood Urea (B. Urea), Creatinine, and body mass index (BMI) were the independent variables, while BP was the dependent variable. The ordinary least squares (OLS) method was first used and evaluated at a significance level of $\alpha = 0.01$. The statistically significant predictive variables in the OLS model were B. Urea and BMI, with a coefficient of determination ($R^2 = 0.60$) for the dependent variable. Furthermore, the multicollinearity test was performed, descriptive statistics for the variables were determined, and the normality of the data was tested. Finally, polynomial regression and Ridge regression were used to control the problem of high multicollinearity resulting from polynomial parts in third-order polynomial models. The main factor affecting BP variability in this method is B. Urea. We conclude, from comparing the results, that the polynomial method is more suitable, with an increase in the value of the coefficient of determination by ($R^2 = 0.70$) and the significance of all combinations between two variables. Furthermore, the higher the degree of the polynomial model, the more complex and more difficult it is to interpret the results.

Index Terms: Blood Pressure, Ordinary Least Squares Regression, Polynomial Regression, Non-linear Modeling, Quadratic Terms

1. INTRODUCTION

Regression analysis is a statistical technique for analyzing the relationship between a dependent variable and one or more independent variables. It helps to explain the relationship between changes in predictor variables and changes in the outcome variable, commonly used for prediction and clarification. Although a straight-line relationship between variables is used in linear regression, real-world data may not always support this assumption [1]. Real-world data frequently shows intricate and curved patterns that call for the application of more advanced non-linear approaches in order to produce reliable forecasts. Rigidity in capturing non-linear relations is the term used to describe the serious miscalculations and poor model performance that can result from relying solely on linear assumptions. Consequently, it is crucial to investigate more trustworthy options that may faithfully replicate the fundamental structure of intricate information [2]. Polynomial regression extends on linear regression and allows the model to recognize more complex and curved patterns in the data. Furthermore, Polynomial regression can be applied to non-linear relationships. By employing higher-degree terms of the independent variables, the correlation coefficients are usually calculated using ordinary least squares (OLS), which minimizes the variance between real and predicted values. While polynomial regression makes the model more flexible, overfitting

Access this article online	
DOI:10.21928/uhdjst.v10n1y2026.pp143-150	**E-ISSN:** 2521-4217 **P-ISSN:** 2521-4209

Corresponding author's e-mail: Zhwan Rashid Aziz, Department of Statistical and Information, College of Administration and Economic, Sulaimani University, Sulaymaniyah, Iraq. E-mail: zhwan.aziz@univsul.edu.iq

Received: 25-02-2026 Accepted: 26-04-2026 Published: 14-05-2026

may be reduced by choosing the right polynomial degree. Polynomial regression can increase model accuracy and offer better knowledge of non-linear relationships when used effectively [3].

1.1. Aim of the Study

The aim of this study is to detect the impact of the independent variables (blood urea [B. Urea], creatinine, and body mass index [BMI]) on dependent variables, blood pressure (BP), by using a statistical model (Polynomial model). It is also to realize which factor in the independent variable has more effect on the dependent variable, as well as to compare between (OLS) and (Polynomial) model estimation to know which of them is a suitable model for estimation.

2. LITERATURE REVIEW

Several researchers have explored the application of polynomial regression and other advanced modeling techniques across various scientific domains. For instance, Kim and Oh (2021) addressed the challenge of predicting strength degradation in fiber-reinforced polymers. They compared artificial neural networks (ANN) with traditional statistical models, including polynomial regression. Their findings demonstrated that while polynomial regression is a reliable baseline, ANN offered superior accuracy in complex alkaline environments [4]. In a different context, Shi *et al.* (2023) focused on improving engineering outcome forecasts by proposing ensemble regression methods. They integrated decision trees with polynomial regression to create the Polynomial Random Forest (PRF) model. Their results indicated that this hybrid approach provided more robustness and precision than traditional polynomial models alone [5]. Furthermore, Manzhos and Ihara (2024) investigated the limitations of kernel methods, such as kernel ridge regression, for sparse data. Their study revealed that in certain high-dimensional scenarios, advanced kernel methods often behave similarly to low-order polynomial regression, emphasizing the continued relevance of polynomial models in medium-sized molecular analysis [6]. Regarding renewable energy optimization, Islam *et al.* (2024) utilized multivariate polynomial regression to analyze a hybrid wind-solar-hydrogen system. By applying this technique within Hybrid Optimization Model for Multiple Energy Resources (HOMER) PRO simulations, they determined that wind resources are the most critical factors influencing energy costs, leading to a more sustainable and cost-effective system design [7]. Finally, the relevant point made by Fu *et al.* (2025) in their discussion is that researchers need to prioritize the methodological rigor while using tools such as polynomial regression. They specifically referred to the traditional block variable methods applied in mediation analysis and suggested a “disaggregated” method as a solution. The R Shiny application that supports their research makes it possible to perform non-linear modeling tools with greater transparency and reiteration of the results [8].

2.1. Methodology (Theoretical Part)

2.1.1. Regression techniques and model specification

The present research focuses on the link between biochemical markers and BP, which is examined through a stepwise procedure that includes both OLS and polynomial regression. At first, the model of multiple linear regression is formulated on the basis of the classic expression for a dependent variable

$$Y = X\beta + \epsilon$$

OLS estimates the model parameters by minimizing the sum of squared residuals [9]. In order to include non-linear relationships, the model is expanded by using a third-degree (cubic) polynomial functional form. This shifts the analysis from a linear framework to a cubic polynomial form representation. The capacity of the model to identify interaction, quadratic, and cubic effects among variables is made possible by this fact [10]. As a result, the structured equation is presented as follows:

$$Y_i = \beta_0 + \sum\nolimits_{j=1}^{k} \beta j X ij + \sum\nolimits_{j=1}^{k} \beta_{jj} X_{ij}^{2} + \sum\nolimits_{j<k} \beta_{jk} X_{ij} X_{ik} + \epsilon_i \qquad (2.1.1)$$

2.1.2. Regularization and diagnostics

The presence of multicollinearity in the polynomial model is indicated by the variance inflation factor (VIF) values exceeding 100 due to polynomial expansion. This necessitates the use of ridge regression (L2 Regularization) to stabilize coefficient estimation. The Ridge loss function is a regularization approach that introduces a tuning parameter (λ) which penalizes large coefficients, thus reducing overfitting [11]:

$$\hat{\beta}_{Ridge} = (X'X + \lambda I)^{-1} X'Y \qquad (2.1.2)$$

To verify the model’s assumptions, the Shapiro–Wilk test is utilized for residual normality and the Durbin-Watson test for autocorrelation [12].

2.1.3. Evaluation criteria

Model performance is assessed using R-squared (R^2) and adjusted R^2 to evaluate the goodness-of-fit. To ensure predictive precision and achieve model parsimony, the root mean square error (RMSE), the Akaike information criterion (AIC), and the Bayesian information criterion are utilized. These metrics ensure that the model generalizes effectively to the population and avoids overfitting [13].

2.2. Data Description

In this study, 180 patient's observations from the Sulaimani Central Laboratory for the years (2023–2024) make up the dataset. The association between the dependent variable, BP (Y), and three continuous independent variables, B. Urea (X_1), creatinine (X_2), and BMI (X_3), is examined in this cross-sectional study. Variable distributions were observed using descriptive statistics, such as mean, standard deviation (SD), and range; the Shapiro–Wilk test ($\alpha = 0.01$) verified that B. Urea and BMI differed significantly from normality. Z-score normalization was used to standardize all independent variables in order to guarantee homogeneity and enhance the stability of the ensuing polynomial regression. These variables are X_1: B. Urea, X_2: Creatinine (Creatin.), X_3: BMI, and (Y): BP: Table 1 below shows the real Data set.

2.3. Statistical Analysis

2.3.1. Descriptive statistics

Table 2 below provides a summary of the descriptive statistics for the variables used in this investigation. These statistics include the mean, SD, and range (minimum and maximum values), give an overall view of the dataset's central tendency and variability.

TABLE 1: Represents the real data set

n	Y	X_1	X_2	X_3
1	100	135.55	9.69	19.82
2	110	144.66	7	28.8
3	80	45.39	2.7	20.16
4	130	76.28	7.06	32.72
5	140	30.8	4.83	28.96
…	…	…	…	…
177	80	24.47	8.3	19
178	150	174.37	5.1	37.82
179	120	149.35	5.1	22.8
180	160	193.06	4.9	20.16

TABLE 2: Summary of descriptive statistics for blood pressure and biochemical variables

Variable	Mean	SD	Minimum	Maximum
BP	130	13.86	80.00	146.67
B. Urea	118.59	60.99	12.10	383.88
Creatinine	7.77	5.28	1.30	42.00
BMI	27.98	7.08	10.96	42.97

BP: Blood pressure, BMI: Body mass index, SD: Standard deviation, B. Urea: Blood urea

According to the table, the sample's mean BP was (130 with SD = 13.86). In terms of the biochemical predictors, the mean B. Urea and Creatinine levels were (118.59, SD = 60.99) and (7.77, SD = 5.28), respectively, while the average BMI level was (27.98, SD = 7.08). Although Creatinine measurements showed a rather small SD (5.28), the statistical evaluation displayed is not a significant predictor in this model. Hence, the above statement shows that the values for B. Urea and BMI are the most significant factors in this study, which shows their reliability as predictors.

2.4. Multiple Linear Regression (OLS) Stage

2.4.1. Initial multicollinearity test (VIF)

The assumption of "No Perfect Collinearity" between the independent variables must be verified to ensure the credibility of the OLS estimator. High correlation between predictor variables is known as multicollinearity, and it can cause unstable results by inflating the variance of coefficient estimations. In this study, the degree of interdependence between B. Urea, Creatinine, and BMI was evaluated using a correlation matrix and the VIF. Table 3 below displays the diagnostic results.

While the original linear model was stable with modest VIF values (varying from 1.036 to 1.306), the diagnostic evaluation using the VIF reveals that the complexity of the third-degree polynomial model generated severe multicollinearity. As high as (74.89) for Creatinine^2 and 41.42 for B. Urea^3, VIF values for a number of interaction terms and higher-order variables were much above the essential threshold of 10 in this advanced stage. A high level of instability in the OLS coefficient estimates is shown by these inflated VIF values. Ridge Regression was used to solve this problem and guarantee an accurate parameter estimate. Ridge Regression successfully stabilized the model and reduced the variance by adding a regularization parameter (λ), offering a more reliable fit for the non-linear association between biochemical markers and BP.

2.4.2. Analysis of variance (ANOVA) and significance

Using an ANOVA, the overall significance of the multiple linear regression model and the individual value of each

TABLE 3: Results of the correlation matrix and VIF

Variable	BP	Blood urea	Creatinine	BMI	VIF
BP	1.000	0.542	0.048	0.745	—
Blood urea	0.542	1.000	0.178	0.456	1.306
Creatinine	0.048	0.178	1.000	0.034	1.036
BMI	0.745	0.456	0.034	1.000	1.267

VIF: Variance inflation factor, BP: Blood pressure, BMI: Body mass index

TABLE 4: Statistical significance of regression coefficients

Variables	Coefficient	SD	t	P>|t|	(0.025–0.975)
Intercept	99.399	0.653	152.305	0.000	98.112–100.688
Blood urea	3.5726	0.746	4.789	0.000	2.100–5.045
Creatinine	−0.261	0.664	−0.394	0.694	−1.572–1.049
Body mass index	8.6802	0.735	11.817	0.000	7.231–10.130

SD: Standard deviationn. Coefficient, SD, t, P>|t|, and the 95% Confidence Interval

TABLE 5: ANOVA table for multiple linear regression

Source	d.f	Sum square	Mean square	F	PR>F
Blood urea	1.0	1758.363	1758.363	22.934	3.541939e-06
Creatinine	1.0	11.893	11.893	0.155	6.941693e-01
Body mass index	1.0	10706.586	10706.586	139.647	4.245927e-24
Residual	176.0	13493.712	76.669	-	-

ANOVA: Analysis of variance, d.f: Degrees of freedom

biochemical marker have been evaluated. Table 5 provides specifics on the findings, including the degrees of freedom, sum of squares, and F-statistics, and whether the factors taken together have a statistically significant impact on BP is determined by this analysis.

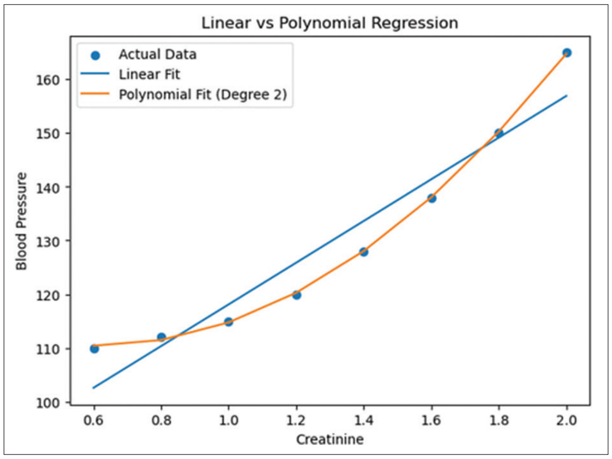

Fig. 1. Comparison between linear and polynomial fits for biochemical markers.

A strong statistical basis for assessing the multiple linear regression model is provided by the ANOVA results shown in Table 5. Several important conclusions are drawn from the analysis: Model significance and predictor influence: The findings show that the model is very significant, especially when it comes to B. Urea and BMI. With an extraordinarily high F-statistic of 139.647 and a *P*-value well below the significance level ($P = 4.24 \times 10^{-24}$), BMI proved to be the most powerful predictor of BP, demonstrating a very strong and dependable linear association. Significance of B. Urea: In addition, B. Urea demonstrated a statistically significant effect on the dependent variable (F = 22.934, $P < 0.001$), confirming its inclusion in this study as a crucial biochemical marker for hypertension prediction. Interestingly, it was discovered that creatinine was not statistically significant in this linear framework (F = 0.155, $P = 0.694$). This implies that although creatinine is an essential biochemical marker, in this particular cohort of 180 cases, its direct linear connection with BP is minimal. Residual analysis and model expansion: The variation that the linear model is unable to account for is reflected in the mean square error, which is represented by the mean square of the residuals (76.669). A simply linear method would not adequately capture the biological complexity of the data, as indicated by the significant residual sum of squares (13493.712) in relation to the predictors. This result offers a compelling scientific argument for switching to polynomial regression (Degree 3) and using Ridge Regression to increase prediction accuracy and capture possible non-linear interactions. The multiple linear regression model can be described as follows using the estimated coefficients from the OLS framework:

$$\text{BP} = 99.399 + 8.680\,(\text{body mass index}) + 3.572\,(\text{blood urea}) - 0.2616\,(\text{Creatinine})$$

This indicates that BMI is the main predictor (F = 139.65, $P < 0.01$), whereas creatinine is not statistically significant ($P = 0.694$). The switch to more delicate non-linear modeling is justified by this linear baseline.

As noticed in Fig. 1, the polynomial expansion (orange line) better represents the underlying distribution, whereas the linear fit (blue line) is unable to explain the non-linear trend of the data.

2.4.3. General measurement and model fit

With an overall significance of F = 52.74 ($P < 0.001$), the OLS model explained 47.3% of the variance ($R^2 = 0.473$; Adjusted $R^2 = 0.464$). The residuals were not normally distributed (Shapiro-Wilk $P < 0.01$), despite the model's lack of significant autocorrelation (Durbin-Watson = 1.95). The OLS residuals (blue points) in Fig. 2. Fitted plot clearly indicates a pattern, suggesting that the linear model ignores the non-linear structure of the data and is unable to maintain constant variance.

The moderate R^2 and this non-normality show that the linear framework is unable to fully capture the complexity of the biochemical data. In addition, BMI and B. Urea were found to be the most important predictors in this population, whereas creatinine was not significant. To address the non-linear nature of the variables and increase predictive accuracy, these diagnostic results support the switch to a third-degree polynomial model in conjunction with Ridge Regression.

2.5. Polynomial Regression

A third-degree polynomial regression was used to capture the intricate non-linear correlations in the biological data. Higher-order terms (such as X^2, X^3) give flexibility to the model, but they also create structural multicollinearity. VIF in this study significantly increased when a linear model was switched to a cubic one. For example, the VIF for Creatinine^2 and Creatinine^3 reached values as high as 74.89 and 67.50, well beyond the conventional cutoff of 10. Since squared and cubic terms are inherently connected with their linear counterparts, these enhanced VIF values are an expected mathematical outcome of polynomial expansion. In an OLS framework, this high multicollinearity may result in overfitting and unstable coefficient estimations. Consequently, these findings offer a solid scientific basis for using Ridge Regression. Ridge Regression ensures a more dependable and generalizable model by addressing the multicollinearity caused by the third-degree expansion and stabilizing the parameter estimations with the introduction of a regularization parameter (λ).

2.5.1. Resolution via ridge regression

Ridge Regression (L^2 Regularization) was used to address the extreme structural multicollinearity (VIF = 74.89) seen in the third-degree polynomial expansion. By including a penalty term (λ) in the regression equation, Ridge Regression stabilizes the model, in contrast to the OLS framework, which yields unstable coefficient estimates in the presence of significant VIF. This method effectively lowers the standard errors of the coefficients, guaranteeing a more accurate estimation of the influence of predictors such as B. Urea and BMI without removing any variables from the model. The models' comparative performance is shown in Fig. 3, which emphasizes how the Ridge approach (red points) offers a more stable fit along the identity line than the OLS and higher-degree polynomial expansions.

The Ridge Regression model's performance metrics are presented in Table 6:

In comparison to the OLS baseline, the Ridge model significantly improves stability by maintaining a strong

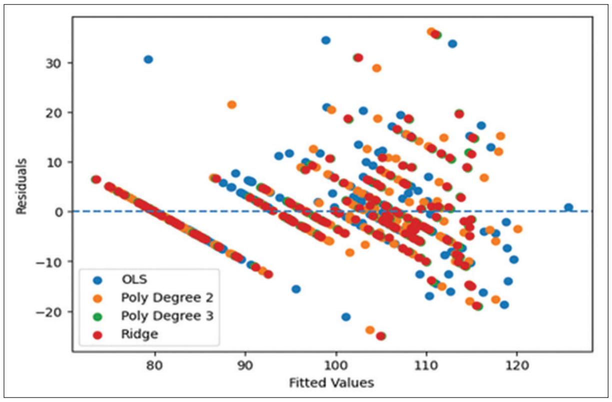

Fig. 2. Residuals versus fitted values for the evaluated models.

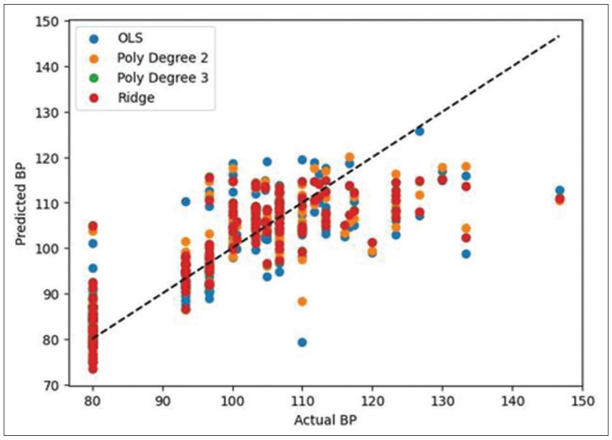

Fig. 3. Actual versus predicted blood pressure values across ordinary least squares, polynomial, and ridge regression models.

TABLE 6: Ridge regression's metrics for performance

Metric	Value
Ridge R^2	0.708
Ridge adjusted R^2	0.673
Ridge RMSE	7.466

RMSE: Root mean square error

TABLE 7: Statistical significance of polynomial regression coefficients (degree 3)

Variables	Coefficient	SD	t	P>|t|	(0.025–0.975)
Constant	102.2104	1.24	82.46	0.0	99.763–104.658
Blood urea	2.2784	1.58	1.442	0.151	−0.841–5.398
Creatinine	1.0664	1.59	0.671	0.503	−2.074–4.206
Body mass index	10.1156	1.32	7.665	0.0	7.509–12.722
Blood urea 2	−2.8296	1.043	−2.712	0.007	−4.89–−0.769
Blood urea creatinine	0.0865	1.746	0.05	0.961	−3.361–3.534
Blood urea body mass index	−1.2663	1.256	−1.008	0.315	−3.747–1.215
Creatinine 2	−0.0981	1.12	−0.088	0.93	−2.31–2.114
Creatinine body mass index	−0.5434	1.385	−0.392	0.695	−3.279–2.192
Body mass index 2	0.7571	0.785	0.964	0.337	−0.794–2.308
Blood urea3	0.6647	0.553	1.202	0.231	−0.428–1.757
Blood urea 2 creatinine	−0.2441	0.52	−0.469	0.64	−1.272–0.783
Blood urea 2 body mass index	0.102	0.802	0.127	0.899	−1.483–1.687
Blood urea creatinine 2	−0.2295	0.787	−0.292	0.771	−1.783–1.324
Blood urea creatinine body mass index	−0.299	1.244	−0.24	0.81	−2.755–2.157
Blood urea body mass index 2	0.8948	0.874	1.023	0.308	−0.832–2.622
Creatinine3	−0.0352	0.167	−0.211	0.833	−0.366–0.295
Creatinine 2 body mass index	0.0007	0.285	0.003	0.998	−0.562–0.564
Creatinine body mass index 2	0.4716	1.544	0.305	0.76	−2.578–3.521
Body mass index 3	−1.4636	0.508	−2.879	0.005	−2.468–−0.46

explanatory power ($R^2 = 0.708$), as seen in Table 5. The model successfully counteracts the variance inflation brought on by the cubic transformations and interactions of biological markers by adding a tiny degree of bias. In the end, this produces a more reliable tool for clinical BP prediction by creating a more generalizable predictive framework that is less susceptible to the strong correlations between higher-order variables.

TABLE 8: ANOVA for polynomial model

Source	Df	Sum square	Mean square	F	PR>F
Blood urea	1.0	374.67	374.67	5.976	0.0156
BMI^2	1.0	382.81	382.81	6.106	0.0145
BMI^{3**}	1.0	519.70	519.70	8.289	0.0045
Residual	160.0	10031.56	62.70	-	-

ANOVA: Analysis of variance, BMI: Body mass index. ** Significant at the 0.01 level.

2.5.2. ANOVA and significance (After multicollinearity correction)

To determine whether the non-linear expansions have distinct predictive value in spite of structural multicollinearity, a secondary ANOVA was employed within the polynomial framework. In order to ascertain if the extra complexity of the third-degree model represents actual biological patterns or just statistical noise, this step is essential.

The results in Table 7 show that the higher-order terms (BMI^2 and BMI^3) have considerable statistical significance ($P < 0.01$). This importance shows that there is a non-linear link between these biochemical indicators and BP, even in spite of the previously mentioned inflated VIF data. This cubic expansion, in contrast to the straightforward OLS framework, captures a substantial explanatory gain, supporting the switch to a more intricate model stabilized by Ridge Regression.

2.5.3. Correlation and multicollinearity analysis

After the polynomial expansion, the multicollinearity analysis showed a notable change in model stability. The addition of third-degree components resulted in significant structural multicollinearity, even though the initial linear OLS framework showed tolerable variance inflation. Higher-order transformations had VIF values as high as 74.89 for (Creatinine^2), while Creatinine^3 also displayed substantial inflation at 67.51. These numbers are far higher than the typical cutoff of 10. In an OLS framework, this severe multicollinearity usually results in unstable coefficient estimations. However, it was not feasible to eliminate items like BMI^3 ($P < 0.01$) in order to lower VIF because doing so would result in the loss of important non-linear information, as the ANOVA results (Table 8) verified their strong statistical significance. In order to stabilize the estimating process, the Ridge Regression technique was utilized. Ridge Regression successfully reduced the effect of these high VIF values by adding a regularization parameter (λ), which allowed the model to become numerically stable while maintaining the crucial non-linear insights. This method effectively resolves the instability present in the high-order polynomial expansion and guarantees that the strong predictive capacity (Ridge $R^2 = 0.708$) is robust and generalizable.

TABLE 9: Comparison of regression model performance

Model	Degree	R^2	Adjusted R^2	RMSE	AIC	BIC	Number of predictors
Linear (OLS)	1	0.607	0.601	8.658	1296.0	1309.0	3
Polynomial	2	0.688	0.672	7.718	1267.0	1298.0	9
Polynomial	3	0.708	0.673	7.465	1275.0	1338.0	19
Ridge regression	3	0.708	0.673	7.466	-	-	19

2.6. Comparison of Models

Table 9 provides a thorough comparison between the non-linear expansions (Polynomial and Ridge Regression) and the linear OLS framework. Several performance criteria, including R^2, Adjusted R^2, and RMSE, were assessed to determine the best model for predicting BP.

The switch from a linear to a non-linear framework led to a notable increase in predicted accuracy, as seen in Table 7. The best method was found to be the Ridge Regression (Degree 3) model, which had the lowest RMSE (7.466) and the highest R^2 (0.708). Despite achieving comparable accuracy, the third-degree polynomial model is less trustworthy for clinical generalization due to its high AIC and BIC values, as well as the previously mentioned extreme multicollinearity. Ridge regression, on the other hand, offers a balanced approach; it uses L2 regularization to guarantee numerical stability while maintaining the crucial non-linear insights (such the important BMI^3 effect). As a result, the Ridge model is the most reliable method for predicting BP in this dataset, providing better accuracy than the more straightforward OLS model without the instability brought on by unregularized polynomial expansions.

3. DISCUSSION

The two regression techniques of this study are used to determine the relationship between BP and biochemical indicators using both linear and non-linear regression methods. The initial OLS analysis found that B. Urea and body mass function are the most significant predictors of BP, displaying a substantial positive correlation (0.745 and 0.542), respectively. This implies that the BP measurements correlate with two variables. However, the linear model's limitations were shown by the diagnostic tests. The non-normal distribution of residuals showed that the linear method could not correctly reflect the complexity of the data, and the high VIF values indicated significant multicollinearity. By moving to polynomial regression, the model's explanatory power increased as indicated by the reduction in RMSE. In this investigation, Ridge Regression proved important. It successfully stabilized the model by treating the multicollinearity produced by the polynomial terms. This illustrates that biochemical connections are frequently non-linear and that advanced regularization methods are required to deliver trustworthy therapeutic insights without running the danger of overfitting.

4. CONCLUSION

In this study, both polynomial regression analysis and OLS analysis were used for studying factors affecting the BP. In this work, the data were collected from 180 patients in a laboratory center at a hospital in Sulaimani city. The BP, from the patients (as dependent variables) and some independent variables (B. Urea, Creatinine, and BMI) were taken. In Table 4, the initial OLS analysis found two variables (B. Urea and BMI) as the most significant predictors of BP, because the (*P*-value) is smaller than the level of significance (alpha = 0.01), and the confidence interval is (=0.148–0.382) for B.Urea and (=0.443–0.687) for BMI, respectively. This means that the BMI and B. Urea were found to be the most important predictors in this population, with (R^2 = 0.60). Furthermore, in Table 7, we conclude, from the comparison of the results, that the polynomial method is a more suitable method the increasing value of (R^2 = 0.70) for all the combinations between two variables that are significant, and the higher the degree of polynomial model is, the more complex and unclear it will be.

REFERENCES

[1] A. O. Sykes. "*An Introduction to Regression Analysis*". Coase-Sandor Institute for Law & Economics Working Paper No. 20, 1993.

[2] A. J. Jones. "New tools in non-linear modelling and prediction". *Computational Management Science*, Vol. 1, no. 2, pp. 109-149, 2004.

[3] S. Kumar. "Polynomial regression". In: *Python for Accounting and Finance*. Palgrave Macmillan, Cham, pp. 445-460, 2024.

[4] Y. Kim and H. Oh. "Comparison between multiple regression analysis, polynomial regression analysis, and an artificial neural network for tensile strength prediction of BFRP and GFRP". *Materials*, vol. 14, no. 17, p. 4861, 2021.

[5] M. Shi, W. Hu, M. Li, J. Zhang, X. Song and W. Sun. "Ensemble regression based on polynomial regression-based decision tree and its application in the *in-situ* data of tunnel boring

machine”. *Mechanical Systems and Signal Processing*, vol. 188, p. 110022, 2023.

[6] S. Manzhos and M. Ihara. “Degeneration of kernel regression with Matern kernels into low-order polynomial regression in high dimension”. *The Journal of Chemical Physics*, vol. 160, no. 2, p. 021101, 2024.

[7] M. K. Islam, N. M. S. Hassan, M. G. Rasul, K. Emami and A. A. Chowdhury. “An off-grid hybrid renewable energy solution in remote Doomadgee of Far North Queensland, Australia”. *Renewable Energy*, vol. 231, p. 120991, 2024.

[8] S. Q. Fu, N. Dimotakis and J. Koopman. “Mediation testing with polynomial regression: A critical review of extant approaches and a researcher’s toolkit for the future”. *Journal of Applied Psychology*, Vol. 111, pp. 1-17, 2025.

[9] S. H. Brown. “Multiple linear regression analysis: A matrix approach with MATLAB”. *Alabama Journal of Mathematics*, vol. 34, p. 1-3, 2009.

[10] B. Meuleman, G. Loosveldt and V. Emonds. “Regression analysis: Assumptions and diagnostics”. In: *Regression Analysis and Causal Inference*, pp. 83-110, 2015.

[11] S. Calonico, M. D. Cattaneo and M. H. Farrell. “Coverage error optimal confidence intervals for local polynomial regression”. *Bernoulli*, vol. 28, no. 4, pp. 2998-3022, 2022.

[12] P. Royston. “Approximating the Shapiro-Wilk W-test for non-normality”. *Statistics and Computing*, vol. 2, no. 3, pp. 117-119, 1992.

[13] R. A. Mahmoud and E. S. Elwakil. “Out-of-step detection for synchronous generators using electrical power analysis and Durbin Watson testing”. *Scientific Reports*, vol. 15, no. 1, p. 20113, 2025.