





## Research Article

# Modeling the Ranked Antenatal Care Visits Using Optimized Partial Least Square Regression

Maryam Sadiq <sup>1</sup>, Alanazi Talal Abdulrahman <sup>2</sup>, Randa Alharbi <sup>3</sup>,  
Dalia Kamal Fathi Alnagar <sup>3,4</sup> and Syed Masroor Anwar <sup>1</sup>

<sup>1</sup>Department of Statistics, University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan

<sup>2</sup>Department of Mathematics, College of Science, University of Hail, Saudi Arabia

<sup>3</sup>Department of Statistics, University of Tabuk, Saudi Arabia

<sup>4</sup>Department of Statistics, Omdurman Islamic University, Sudan

Correspondence should be addressed to Maryam Sadiq; [hussainulahmad@gmail.com](mailto:hussainulahmad@gmail.com)

Received 22 December 2021; Revised 9 February 2022; Accepted 11 February 2022; Published 14 March 2022

Academic Editor: Plácido R. Pinheiro

Copyright © 2022 Maryam Sadiq et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The frequency and timing of antenatal care visits are observed to be the significant factors of infant and maternal morbidity and mortality. The present research is conducted to determine the risk factors of reduced antenatal care visits using an optimized partial least square regression model. A data set collected during 2017-2018 by Pakistan Demographic and Health Surveys is used for modeling purposes. The partial least square regression model coupled with rank correlation measures are introduced for improved performance to address ranked response. The proposed models included  $PLS\rho_s$ ,  $PLS\tau_A$ ,  $PLS\tau_B$ ,  $PLS\tau_C$ ,  $PLS_D$ ,  $PLS\tau_{GK}$ ,  $PLS_G$ , and  $PLS_U$ . Three filter-based factor selection methods are executed, and leave-one-out cross-validation by linear discriminant analysis is measured on predicted scores of all models. Finally, the Monte Carlo simulation method with 10 iterations of repeated sampling for optimization of validation performance is applied to select the optimum model. The standard and proposed models are executed over simulated and real data sets for efficiency comparison. The  $PLS\rho_s$  is found to be the most appropriate proposed method to model the observed ranked data set of antenatal care visits based on validation performance. The optimal model selected 29 influential factors of inadequate use of antenatal care. The important factors of reduced antenatal care visits included women's educational status, wealth index, total children ever born, husband's education level, domestic violence, and history of cesarean section. The findings recommended that partial least square regression algorithms coupled with rank correlation coefficients provide more efficient estimates of ranked data in the presence of multicollinearity.

## 1. Introduction

Pakistan sets targets to minimize the maternal mortality ratio (MMR) to 140 per 100,000 live births by 2015 by increasing skilled birth attendants and improving access to reproductive health care as the fifth Millennium Development Goal (MDG) for improving maternal health suggested. The MDG progress assessment reported that Pakistan was not close to attaining the target in 2015. Recently, Pakistan has endorsed the Sustainable Development Goals (SDGs), committing to decrease the MMR to 70 per 100,000 live births by 2030 by increasing skilled birth attendance, facilitation to modern contraception, and

extending coverage of health workers. The Government of Pakistan took initiatives and made good progress in maternal health indicators during the last decade, and a significant decline was reported in MMR from 276 to 178 [1, 2]. Pregnancy-related morbidity and mortality can be reduced by improving access and facilitation to maternal health care services. At least four antenatal care visits (ANC) are recommended to skilled personnel to avoid any pregnancy-related complication [3]. Nearly 12% Pakistani women reported no ANC throughout their pregnancy, 36% have less than four visits, and 52% claimed four or more visits [1]. Several studies have assessed the significant influential factors of antenatal care

attendance in Pakistan without considering the frequency of ANC [4, 5]. Poisson regression, negative binomial regression, zero-inflated, and hurdle regression models have been commonly used to model the count of ANC visits [6, 7]. Binary logistic regression and a multinomial logistic regression model are also found to study the use and ranks of ANC visits [5, 8]. Advancements in health research generate public health data having many covariates, where some or all may be correlated. Several studies have been conducted to identify influential factors of different public health concerns using multiple statistical tools and techniques [9–13]. The partial least square (PLS) regression model has been the concern of interest as a statistical method for modeling data having multicollinearity during the last few decades. A variety of modified PLS algorithms have been introduced for superior model performance [14]. Most PLS algorithms model continuous factors, and a few are specifically designed for categorical framework but no specific algorithm is projected to address the ranked data. To fill the gap of obtaining the optimal model for the ranked response variable, modified PLS algorithms based on ranked correlation loading weights are introduced. The main motivation of the present study is to propose the modified PLS algorithms to particularly address the ranked response factor in the presence of multicollinearity. To improve the PLS regression model, eight algorithms based on rank correlation measures including Spearman's rank correlation coefficient, Kendall's  $\tau_A$  rank correlation coefficient, Kendall's  $\tau_B$  rank correlation coefficient, the Stuart-Kendall  $\tau_C$  rank correlation coefficient, Somers' delta ( $D$ ), Goodman-Kruskal's tau  $\tau_{GK}$ , Goodman-Kruskal's gamma ( $G$ ), and Thiel's  $U$  correlation coefficient are proposed in this study. To the best of our knowledge, no previous research has considered multicollinear covariates in modeling the ranking of ANC visits of Pakistani women. Thus, the objectives of this study are twofold: (i) to develop a regression model for the ranked response covering the issue of multicollinearity and (ii) to determine the risk factors for inadequate use of antenatal care. This study introduced eight novel PLS algorithms addressing the concern of multicollinearity for a ranked response which is never discussed earlier. The proposed and standard algorithms are executed on a real-life application of ANC data for comparison purposes. These algorithms will facilitate users to obtain more efficient models than the standard PLS approaches for specifically ranked data. Regarding the clinical importance of this study, the influential selected variables of ANC will help maximize the chances for a normal pregnancy by providing priority interventions, increasing coverage, and improving health quality. The novel contribution of this study included:

- (i) eight new PLS algorithms based on rank correlation loading weights are proposed to address rank response
- (ii) the significant factors of ANC utilization of Pakistani women are identified

## 2. Methodology

**2.1. Data Set.** The data acquired from Pakistan demographic and health survey (PDHS) 2017-2018 for Baluchistan is used

for the present study. A total of 943 observations (women) with 43 factors are included in the analysis. The frequency of ANC is considered the ordinal response factor ( $y$ ) which is ranked as inadequate (0-3 visits), intermediate (4-7 visits), and adequate (8-9 visits).

**2.2. Partial Least Square Regression (PLSR).** Consider the regression model  $y = \alpha + Z\beta + \varepsilon$ , where  $\alpha$  and  $\beta$  are the unknown regression parameters and  $\varepsilon$  is the error term. Let  $Z_{(n,p)}$  is the matrix of explanatory variables and is assumed to be linearly related with the response  $y_{(n,1)}$  and suppose some  $C$  (where  $C \leq p$ ) to represent the number of components for prediction. Then, for  $c = 1, \dots, C$ , the general algorithm executes as

- (1) The loading weights are  $w_c = Z'_{c-1}y_{c-1}$
- (2) The score vector is  $t_c = Z_{c-1}w_c$
- (3) Evaluate  $Z$ -loadings ( $p_c$ ) and  $Y$ -loadings ( $q_c$ ) by  $p_c = Z'_{c-1}t_c/t'_c t_c$  and  $q_c = y'_{c-1}t_c/t'_c t_c$ , respectively
- (4) Deflate  $Z_{c-1}$  and  $y_{c-1}$  by  $Z_c = Z_{c-1} - t_c p'_c$  and  $y_c = y_{c-1} - t_c q_c$
- (5) Repeat the algorithm, if  $c < C$

Consider that  $W$ ,  $S$ ,  $P$ , and  $q$  are the matrices/vectors to compile the loading weights, scores,  $Z$ -loadings, and  $Y$ -loadings computed at each iteration of the algorithm, respectively. The regression estimators of the PLSR model are computed by  $\hat{\beta} = W(P'W)^{-1}q$  and  $\hat{\alpha} = \bar{y} - \bar{Z}\hat{\beta}$  [15]. The general steps of standard PLSR are presented in Figure 1.

The standard PLS is designed for continuous dependent variable  $y$  but if the response is measured on a rank scale then this standard method may not work well. The most important phase of the PLS algorithm is to compute loading weight having the ability to choose significant factors. Loading weights compute the correlation between the dependent variable and predictors. If the data set is ranked then Spearman's rank correlation coefficient, Kendall's  $\tau_A$  rank correlation coefficient, Kendall's  $\tau_B$  rank correlation coefficient, the Stuart-Kendall  $\tau_C$  rank correlation coefficient, Somers' delta ( $d$ ), Goodman-Kruskal's tau  $\tau_{GK}$ , Goodman-Kruskal's Gamma ( $G$ ), and Thiel's  $U$  correlation coefficient are the recommended measures of rank correlation. These measures of association are used to compute the loading weights of the PLS algorithm. The modified loading weights of PLSR are visually displayed in Figure 2.

**2.2.1. PLS $\rho_s$ .** Spearman's rank correlation coefficient or Spearman's  $\rho$  ( $\rho_s$ ) [16] is a nonparametric measure of rank correlation using a monotonic function. It is used to compute the weights of as

$$w_{\rho_s} = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}, \quad (1)$$

where  $d_i$  denotes the difference between the two ranks of

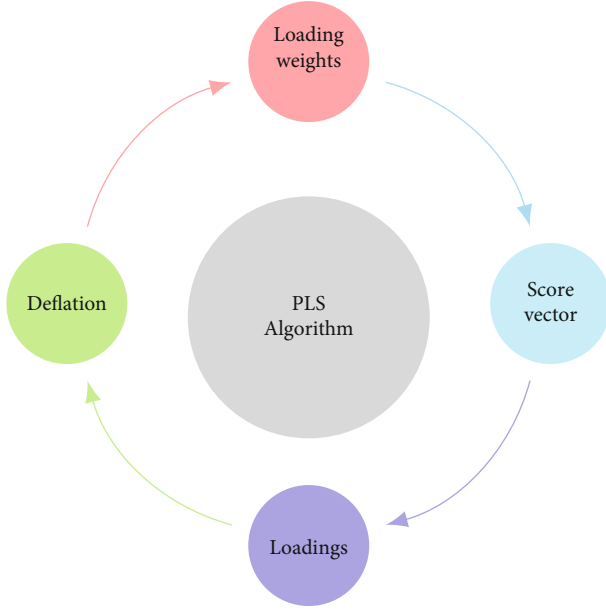


FIGURE 1: Standard PLS algorithm.

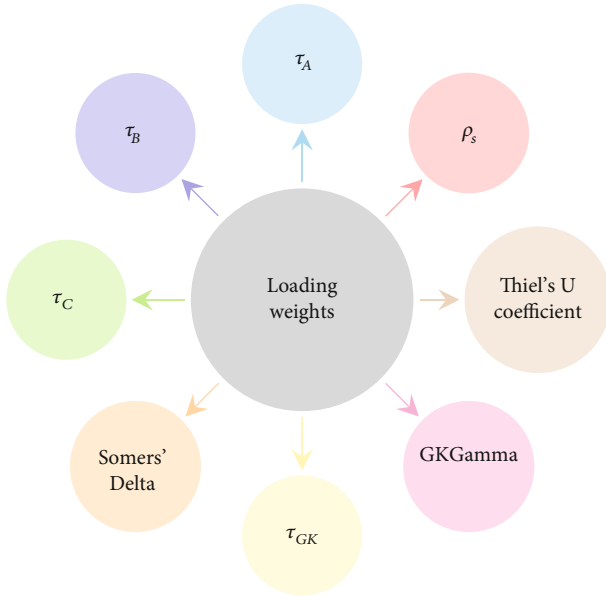


FIGURE 2: Modified loading weights of PLS.

each observation and  $n$  is the number of observations and the modified PLSR algorithm is referred to as  $PLS\rho_s$ .

2.2.2.  $PLS\tau_A$ . The Kendall rank correlation coefficient or Kendall's  $\tau$  coefficient is a measure of rank correlation. The tau-A ( $\tau_A$ ) will not make any adjustment for ties [17]. It is used to define the PLS loading weights as

$$w_{\tau_A} = \frac{n_c - n_d}{n_0}, \quad (2)$$

where  $n_c$  is the number of concordant pairs,  $n_d$  is the number of discordant pairs, and  $n_0 = n(n-1)/2$ , and the modified algorithm is named as  $PLS\tau_A$ .

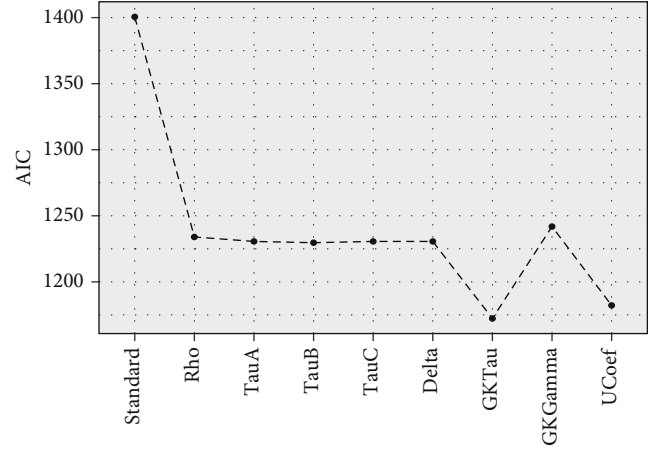


FIGURE 3: The AIC performance of PLSR models based on rank correlation coefficients' oversimulated data without considering any variable selection method is presented.

2.2.3.  $PLS\tau_B$ . Kendall's tau-B makes adjustments for ties [18]. The PLS loading weights are altered by using  $\tau_B$  as

$$w_{\tau_B} = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}}, \quad (3)$$

where  $n_1 = \sum_i t_i(t_i - 1)$ ,  $n_2 = \sum_j u_j(u_j - 1)$ ,  $t_i$  is the number of tied value in the  $i^{\text{th}}$  group of ties and  $u_j$  is the number of tied value in the  $j^{\text{th}}$  group of ties and the proposed model is termed as  $PLS\tau_B$ .

2.2.4.  $PLS\tau_C$ . The Stuart-Kendall (tau-c) is more suitable for contingency tables [19]. The  $\tau_C$  replaced the weights of PLS as follows:

$$w_{\tau_C} = \frac{2(n_c - n_d)}{n^2((m-1)/m)} \quad (4)$$

where  $m$  is the minimum number among rows and columns, and the modified PLSR algorithm is called  $PLS\tau_C$ .

2.2.5.  $PLS\tau_D$ . The PLS loading weights based on Somers' delta(D) [20] of variable  $Y$  with respect to variable  $Z$  are defined as

$$w_D = \frac{\tau(Z, Y)}{\tau(Z, Z)}. \quad (5)$$

Kendall's tau  $\tau$  is symmetric, whereas Somers'  $D$  is asymmetric in  $Z$  and  $Y$ , and the model is named as  $PLS_D$ .

2.2.6.  $PLS\tau_{GK}$ . Goodman-Kruskal's tau  $\tau_{GK}$  [21] is integrated as PLS loading weights as

$$w_{\tau_{GK}} = \frac{2(n_c - n_d)}{n_0}, \quad (6)$$

and the modified algorithm is called  $PLS\tau_{GK}$ .

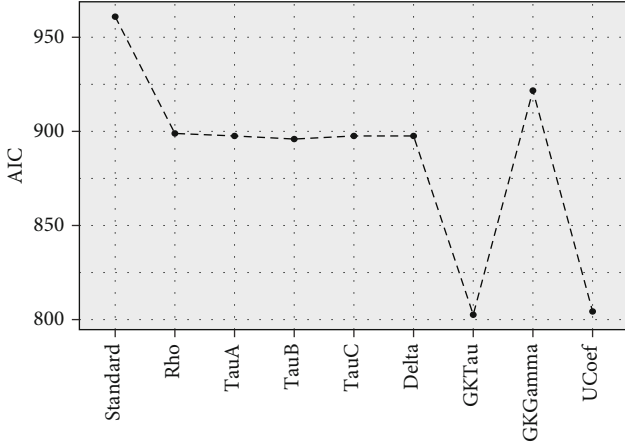


FIGURE 4: The AIC performance of PLSR models based on rank correlation coefficients' oversimulated data integrated with LW variable selection method is presented.

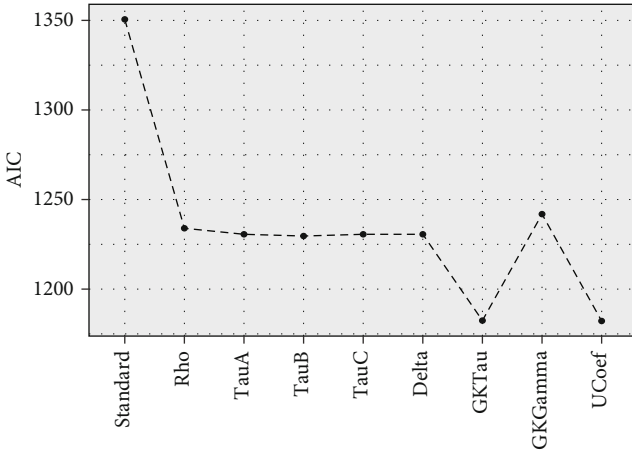


FIGURE 5: The AIC performance of PLSR models based on rank correlation coefficients' oversimulated data integrated with SMC variable selection method is presented.

2.2.7.  $PLS_G$ . The estimate of Goodman-Kruskal's gamma ( $G$ ) [21] is used as loading weights of PLS

$$w_G = \frac{n_c - n_r}{n_c + n_r}, \quad (7)$$

where  $n_c$  is the number of concordant pairs and  $n_r$  is the number of reversed pairs. Goodman-Kruskal's gamma drop ties, and the PLSR model is named as  $PLS_G$ .

2.2.8.  $PLS_U$ . Thiel's  $U$  correlation coefficient or uncertainty coefficient [22] altered the PLS loading weights as

$$w_U = \frac{H(Z) - H(Z|Y)}{H(Z)}, \quad (8)$$

where  $H(Z)$  represents the entropy of a single distribution and  $H(Z|Y)$  represents the conditional entropy, and the modified PLSR algorithm is referred to as  $PLS_U$ .

2.3. Filter-Based Factor Selection Methods. Several variable selection methods integrated with PLSR have been introduced. The following are considered here.

2.3.1. The Loading Weight ( $LW$ ). The loading weights  $w_j$  are used to measure of importance of predictors and are defined as [23].

$$LW = \left| \frac{w_{a,j}}{\max(w_a)} \right|. \quad (9)$$

2.3.2. The Regression Coefficients ( $RC$ ). The regression coefficient estimates are computed as [23]

$$RC = W(P'W)^{-1}q. \quad (10)$$

2.3.3. Significance Multivariate Correlation ( $SMC$ ). The significance multivariate correlation measure is used to reduce the effect of irrelevant predictors and enhance the influence of significant variables included in the model. The SMC [23] is computed as

$$SMC = \frac{\text{MeanSquareRegression}}{\text{MeanSquareResidual}}. \quad (11)$$

### 3. Results

Initially, the PLS models with modified loading weights are executed for simulated data set for ranked variables. A sample of size 1000 with 100 predictors is generated. The response variable and 50% predictors are generated over 3 ranks, and the remaining explanatory variables are distributed over 4 ranks.

Spearman's coefficient, Kendall's coefficient-A, Kendall's coefficient-B, Stuart-Kendall's, Somers' delta, Goodman-Kruskal's tau, Goodman-Kruskal's gamma, and Thiel's  $U$  coefficient are used as loading weights of the PLS algorithm to fit oversimulated data set to observe the variation in performance of standard and proposed models based on Akaike information criterion (AIC).

Figure 3 showed the efficiency of models established by AIC and indicated that PLSR algorithms with modified loading weights have higher efficiency (lower mean AIC) compared to standard PLSR for a ranked response. The PLSR model with  $\tau_{GK}$  as modified loading weight showed optimum performance compared to eight other models without integrating any variable selection method. All other proposed models also evidenced higher efficiency compared to standard PLSR. Figures 4 and 5 also demonstrated the higher accuracy of proposed models compared to standard PLSR algorithm integrated with LW and SMC variable selection methods. Both figures depicted that  $PLS_{\tau_{GK}}$  and  $PLS_U$  have optimum performance compared to all other models. The standard and modified models are also executed over the real data set of ANC for comparison of accuracy. The data set of ANC visits had 43 predictors sampled over 943 samples (mothers). The Spearman rank correlation coefficient is used to examine the multicollinearity in the data.

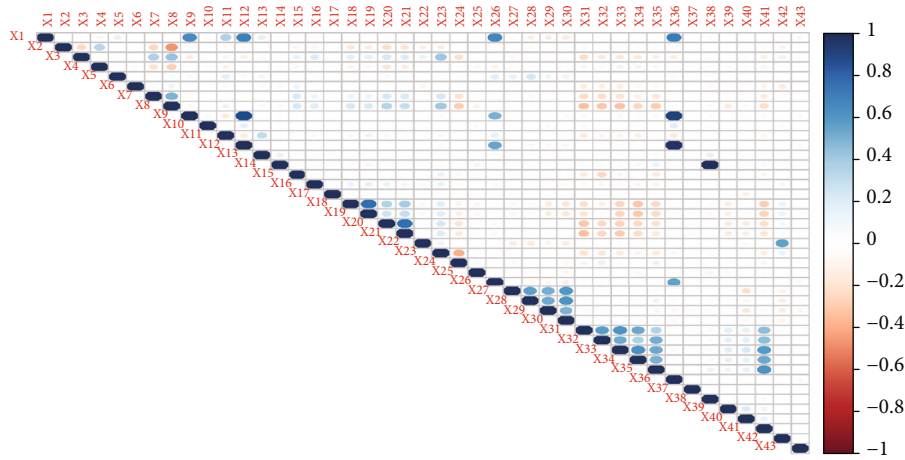


FIGURE 6: Correlogram by Spearman rank correlation matrix is presented. Color intensity and the size of the circle are proportional to the strength of the correlation among covariates.

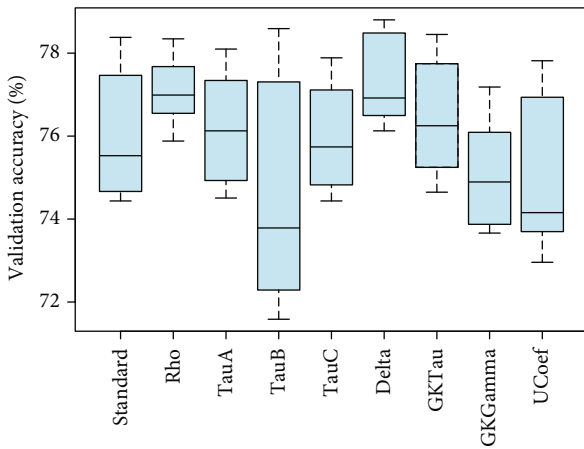


FIGURE 7: The validation accuracy of PLSR models based on rank correlation coefficients including Spearman’s correlation coefficient (Rho), Kendall coefficient-A (tau-A), Kendall coefficient-B (tau-B), Stuart-Kendall (tau-C), Somers’ delta (Delta), Goodman-Kruskal’s tau (GKTau), Goodman-Kruskal’s gamma (GKGamma), and Thiel’s  $U$  coefficient (UCoef) models is presented.

The correlogram map measured strong correlation among 16 covariates while intermediate correlation among several other predictors is observed and shown in Figure 6. The existence of multicollinearity recommends the applicability of PLSR to deal ranked data with multicollinearity.

The frequency of ANC is classified into three ranks as inadequate, intermediate, and adequate. The ratio of 70 : 30 is used to randomly split data into training and testing sets, respectively. Initially, PLSR integrated with rank correlation coefficients as loading weights is executed. The Spearman’s coefficient, Kendall’s coefficient-A, Kendall’s coefficient-B, Stuart-Kendall’s, Somers’ delta, Goodman-Kruskal’s tau, Goodman-Kruskal’s gamma, and Thiel’s  $U$  coefficient are used to modify loading weights. Then, three filter-based factor selection methods are performed on each of the nine PLSR models. Leave-one-out cross-validation by linear discriminant analysis is measured on predicted scores of all

27 models. Finally, the Monte Carlo simulation method is used with 10 iterations of repeated sampling for optimization of validation performance. The standard PLSR is compared with the PLSR integrated with Spearman’s coefficient ( $PLS\rho_s$ ), PLSR integrated with Kendall’s coefficient-A ( $PLS\tau_A$ ), PLSR integrated with Kendall’s coefficient-B ( $PLS\tau_B$ ), PLSR integrated with the Stuart-Kendall ( $PLS\tau_C$ ), PLSR integrated with Somers’ delta ( $PLS_D$ ), PLSR integrated with Goodman-Kruskal’s tau ( $PLS\tau_{GK}$ ), PLSR integrated with Goodman-Kruskal’s gamma ( $PLS_G$ ), and PLSR integrated with Thiel’s  $U$  coefficient ( $PLS_U$ ).

Figure 7 showed the comparison of validation performance of standard PLSR and eight proposed PLSR models integrated with correlation coefficients without considering any variable selection method. These results depicted that  $PLS\rho_s$  and  $PLS_D$  have optimum performance compared to standard PLSR and other proposed PLSR models for the observed data of ANC visits. The  $PLS\tau_A$  and  $PLS\tau_{GK}$  also have relatively higher accuracy than standard PLSR. In Figure 8, the loading weight factor selection method is incorporated with each PLSR model. The inclusion of the variable selection method enhanced the overall performance of standard and modified PLSR models. The results showed that  $PLS\tau_A$  and  $PLS\tau_B$  are more efficient in terms of optimization accuracy compared to standard and modified PLS models. A similar pattern of performance for  $PLS\rho_s$  and  $PLS\tau_C$  is observed compared to standard PLSR. Four other proposed PLSR models integrated with rank correlation coefficients showed slightly lower accuracy compared to standard PLSR.

Figure 9 established performance comparison of the PLSR model based on the RC factor selection method. The results demonstrated that compared to PLSR, the six proposed methods including  $PLS\rho_s$ ,  $PLS\tau_A$ ,  $PLS\tau_B$ ,  $PLS\tau_C$ ,  $PLS_D$ , and  $PLS_U$  featured incremental performance after accounting RC selection method. Two other proposed methods  $PLS\tau_{GK}$  and  $PLS_G$  demonstrated approximately identical efficiency as standard PLSR. Standard and modified PLSR models embedded with the SMC factor selection method are compared in Figure 10 in terms of validation accuracy. The two proposed PLSR models including  $PLS\rho_s$



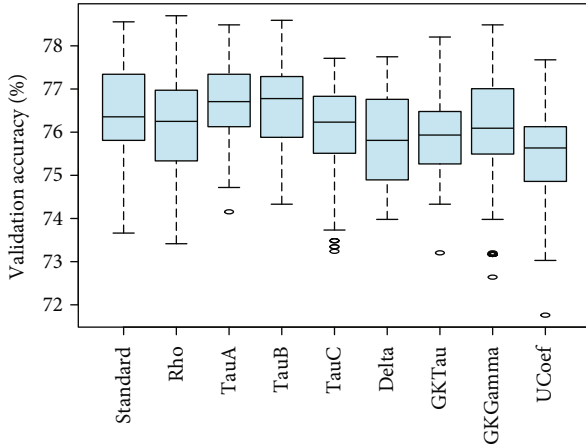


FIGURE 8: The validation accuracy of PLSR models based on rank correlation coefficients integrated with LW factor selection method is presented.

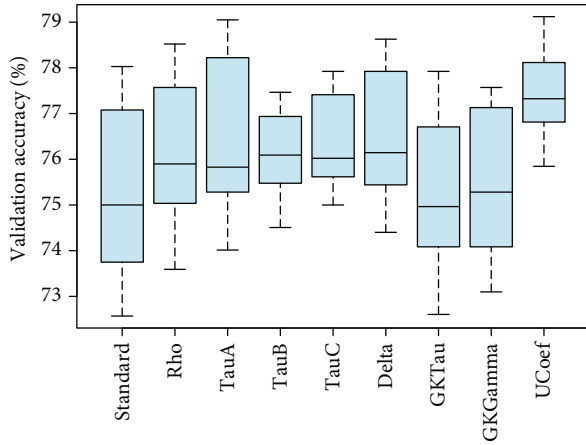


FIGURE 9: The validation accuracy of PLSR models based on rank correlation coefficients integrated with the RC factor selection method is presented.

and  $PLS\tau_C$  evidenced higher accuracy than standard PLSR. The other four modified models including  $PLS\tau_A$ ,  $PLS_D$ ,  $PLS_G$ , and  $PLS_U$  measured analogous accuracy as PLSR.

Comparison based on validation accuracy supported that  $PLS\rho_s$  is found to be the most appropriate proposed method to model the observed ranked data set of ANC. Figure 7 represented the optimal performance of  $PLS\rho_s$  compared to all other models without considering any variable selection method. Integrated with RC factor selection methods, the proposed  $PLS\rho_s$  showed higher efficiency compared to standard PLS visualized in Figure 9. Moreover,  $PLS\rho_s$  established the highest optimization accuracy of nearly 78% among all other methods in 10 combined with the SMC factor selection method. Based on this evidence,  $PLS\rho_s$  featured with the SMC method is finally picked for the selection of influential factors of ANC. For extraction of influential factors of ANC,  $PLS\rho_s$  coupled with SMC is executed and estimates of 29 variables are presented in Table 1 with regression estimates.

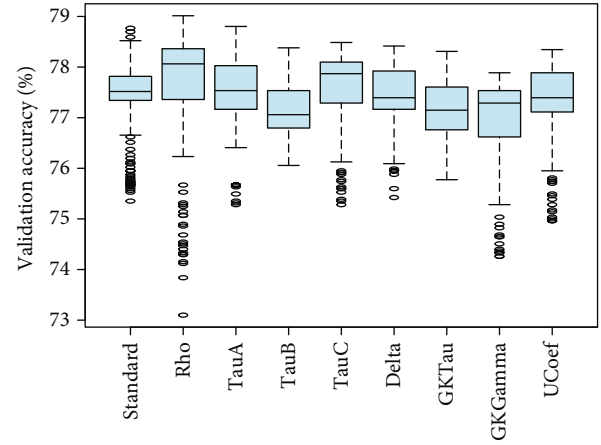


FIGURE 10: The validation accuracy of PLSR models based on rank correlation coefficients integrated with SMC factor selection method is presented.

#### 4. Discussion

To examine the significant predictors associated with ANC, sample data obtained from PDHS (2017-2018) is used. The occurrence of multicollinearity pointed to the application of PLS being a popular substitute for the standard regression model. Data is randomly divided into testing and training sets. Eight PLS algorithms established on rank correlation coefficients are introduced to address particularly the ranked response and compared with the standard PLS model to prove the improved efficiency in model building. The proposed models include Spearman's coefficient ( $PLS\rho_s$ ), PLSR integrated with Kendall's coefficient-A ( $PLS\tau_A$ ), PLSR integrated with Kendall's coefficient-B ( $PLS\tau_B$ ), PLSR integrated with the Stuart-Kendall ( $PLS\tau_C$ ), PLSR integrated with Somers' delta ( $PLS_D$ ), PLSR integrated with Goodman-Kruskal's tau ( $PLS\tau_{GK}$ ), PLSR integrated with Goodman-Kruskal's Gamma ( $PLS_G$ ), and PLSR integrated with Thiel's  $U$  coefficient ( $PLS_U$ ).

Furthermore, three variable selection methods are integrated with standard and modified PLS algorithms to estimate the accuracy to examine the variation in the performance of modified and standard PLS models with and without variable selection methods. The variable selection methods, namely, loading weights, regression coefficients, and significance multivariate correlation are considered here. The validation performance is computed for 10 iterations to examine the efficiency of nine PLS models integrated with variable selection methods.

Comparison based on validation performance supported that  $PLS\rho_s$  is found to be the most appropriate proposed method to model the observed ranked data set of ANC. Figure 7 represented the optimal performance of  $PLS\rho_s$  compared to all other models without considering any variable selection method. Integrated with RC factor selection methods, the proposed  $PLS\rho_s$  showed higher efficiency compared to standard PLS visualized in Figure 9. Moreover,  $PLS\rho_s$  established the highest optimization accuracy of nearly 78% among all other methods in 10 combined with

TABLE 1: Regression coefficient estimates of influential factors for  $PLS\rho_s$  coupled with SMC.

Factor	Coefficient
Age of women	0.010
Type of residence	-0.036
Women's educational level	0.061
Source of drinking water	-0.036
Age of household head	0.023
Frequency of watching television	0.022
Wealth index	0.052
Total children ever born	0.048
Births in last five years	0.016
Number of living children	-0.037
Ever had a terminated pregnancy	-0.034
Visited health facility last 12 months	0.021
Previous birth a caesarean section	0.047
Getting medical help for self: getting permission to go	-0.030
Getting medical help for self: getting money needed for treatment	0.045
Getting medical help for self: distance to health facility	0.021
Heard of tuberculosis	-0.023
Husband's education level	0.040
Respondent's occupation	0.037
Husband's age	-0.017
Beating justified if wife neglects the children	-0.012
Beating justified if wife argues with husband	-0.049
Beating justified if wife refuses to have sex with husband	0.042
Sex of child	0.015
Number of pregnancy losses	0.035
Blood relationship with husband	-0.035
Had a say in choosing the husband	0.026
Ever heard of hepatitis B or C	-0.028
Receive any cash/kind benefit from Benazir Income Support Program	-0.024

the SMC factor selection method. Based on this evidence,  $PLS\rho_s$  featured with the SMC method is finally picked for the selection of influential factors of ANC.

Regarding validation accuracy, very important and interesting facts are observed about the comparison of efficiency for ranked data. Primarily,  $PLS\rho_s$  and  $PLS_D$  have optimum performance compared to standard PLSR and other proposed models without considering any factor selection method. For the observed data set, the  $PLS\tau_A$  and  $PLS\tau_B$  combined with the LW variable selection method are found to be more efficient in terms of optimization accuracy compared to standard and modified PLS models. Integrated with RC method for variable selection, the  $PLS\rho_s$ ,  $PLS\tau_A$ ,  $PLS\tau_B$ ,  $PLS\tau_C$ ,  $PLS_D$ , and  $PLS_U$  featured incremental performance compared to standard PLS. The  $PLS\tau_{GK}$  and  $PLS_G$  are found to exhibit approximately similar efficiency as standard PLSR. The  $PLS\rho_s$  and  $PLS\tau_C$  embedded with the SMC factor selection method evidenced optimum accuracy compared to standard PLS. Considering all validation comparisons, it is noticed that the modified models integrated with rank correlation coefficient exhibit higher efficiency for ranked data set of ANC compared

to the standard PLS algorithm. The  $PLS\rho_s$  coupled with SMC is suggested for modeling the ANC ranked data and 29 influential factors are observed to discriminate the ANC ranks. The proposed algorithms for rank response will facilitate researchers to address the regression models more efficiently even in the presence of multicollinearity in different fields of research. Since the rank response is specifically addressed rarely, the findings of this study offer new, potentially useful information for this ranked population. In the future, these algorithms may be integrated with other variable selection methods to observe the efficiency. Also, the proposed study can be extended for neutrosophic statistics [9]. The main limitation of this study is the small number of predictors as every possible factor was not available for the target population and also the interaction effects are not included.

## 5. Conclusion

Proposed PLS algorithms integrated with rank correlation coefficients are observed to be a better option with regard to model efficiency and variable selection of ranked

simulated and real data sets. This suggests that these rank measure-based PLS algorithms provide models with superior potential. The  $PLS\rho_s$  coupled with SMC identified the significant predictors of ANC using the optimized model for the observed data. The modified PLS models have the ability to address multicollinear ranked data more effectively. Regarding the clinical importance of this study, the influential selected variables of ANC will help maximize the chances for a normal pregnancy by providing priority interventions, increasing coverage, and improving health quality.

## Data Availability

Data is available at <https://dhsprogram.com/data/>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors would like to acknowledge DHS (Demographic and Health Surveys) who made their data available for free.

## References

- [1] NIPS I, *Pakistan demographic and health survey 2017-18*, Islamabad, Pakistan, and Rockville, Maryland, USA: NIPS, ICF, 2019.
- [2] Demographic P, *Health Survey 2012-13*, MA: National Institute of Population Studies and ICF International; 2013, Islamabad and Calverton, 2015, <https://dhsprogram.com/data>.
- [3] Organization WH, *Standards for Maternal and Neonatal Care*, 2007.
- [4] H. Budhwani, K. R. Hearld, and H. Harbison, "Individual and area level factors associated with prenatal, delivery, and postnatal care in Pakistan," *Maternal and child health journal.*, vol. 19, no. 10, pp. 2138–2146, 2015.
- [5] J. W. Noh, K. Ym, L. J. Lee, N. Akram, F. Shahid, and Y. D. Kwon, "Factors associated with the use of antenatal care in Sindh province, Pakistan: a population-based study," *PLoS One*, vol. 14, no. 4, p. e0213987, 2019.
- [6] M. Aslam, M. Sadiq, and T. Mehmood, "Assessment of maternal health services utilization in Pakistan: the role of socio-demographic characteristics," *Asian Biomedicine.*, vol. 14, no. 1, pp. 3–7, 2020.
- [7] K. R. Bhowmik, S. Das, and M. A. Islam, "Modelling the number of antenatal care visits in Bangladesh to determine the risk factors for reduced antenatal care attendance," *PLoS One*, vol. 15, no. 1, article e0228215, 2020.
- [8] S. T. Adedokun and S. Yaya, "Correlates of antenatal care utilization among women of reproductive age in sub-Saharan Africa: evidence from multinomial analysis of demographic and health surveys (2010-2018) from 31 countries," *Archives of Public Health.*, vol. 78, no. 1, pp. 1–10, 2020.
- [9] R. A. K. Sherwani, H. Shakeel, W. B. Awan, M. Faheem, and M. Aslam, "Analysis of COVID-19 data using neutrosophic Kruskal Wallis H test," *BMC Medical Research Methodology.*, vol. 21, no. 1, pp. 1–7, 2021.
- [10] G. S. Rao and M. Aslam, "Inspection plan for COVID-19 patients for Weibull distribution using repetitive sampling under indeterminacy," *BMC Medical Research Methodology.*, vol. 21, no. 1, p. 229, 2021.
- [11] S. Arif, H. Khan, M. Aslam, and M. Farooq, "Factors influencing exclusive breastfeeding duration in Pakistan: a population-based cross-sectional study," *BMC Public Health*, vol. 21, no. 1, pp. 1–10, 2021.
- [12] M. Azam, M. Aslam, J. Basharat, M. A. Mughal, M. S. Nadeem, and F. Anwar, "An empirical study on quality of life and related factors of Pakistani breast cancer survivors," *Scientific Reports*, vol. 11, no. 1, p. 24391, 2021.
- [13] M. Aslam, G. S. Rao, M. Saleem, R. A. K. Sherwani, and C. H. Jun, "Monitoring mortality caused by COVID-19 using gamma-distributed variables based on generalized multiple dependent state sampling," *Computational and Mathematical Methods in Medicine.*, vol. 2021, pp. 1–17, 2021.
- [14] M. Sadiq, T. Mehmood, and M. Aslam, "Identifying the factors associated with cesarean section modeled with categorical correlation coefficients in partial least squares," *PLoS One*, vol. 14, no. 7, article e0219427, 2019.
- [15] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
- [16] J. L. Myers, A. D. Well, and R. F. Lorch Jr., *Research Design and Statistical Analysis*, Routledge, 2013.
- [17] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1-2, pp. 81–93, 1938.
- [18] A. Agresti, *Analysis of Ordinal Categorical Data*, vol. 656, John Wiley & Sons, 2010.
- [19] K. J. Berry, J. E. Johnston, S. Zahran, and P. W. Mielke, "Stuart's tau measure of effect size for ordinal variables: some methodological considerations," *Behavior Research Methods.*, vol. 41, no. 4, pp. 1144–1148, 2009.
- [20] R. H. Somers, "A new asymmetric measure of association for ordinal variables," *American sociological review.*, vol. 27, no. 6, pp. 799–811, 1962.
- [21] L. A. Goodman and W. H. Kruskal, "Measures of association for cross classifications, IV: simplification of asymptotic variances," *Journal of the American Statistical Association.*, vol. 67, no. 338, pp. 415–421, 1972.
- [22] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge university press, 2007.
- [23] T. Mehmood, M. Sadiq, and M. Aslam, "Filter-based factor selection methods in partial least squares regression," *IEEE Access.*, vol. 7, pp. 153499–153508, 2019.