Journal of AL-Rafidain University College for Sciences (2025); Issue 57; 136 -143

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS
Journal of AL-Rafidain University College for Sciences

The Use of Cone Projections and Quadratic programming in Estimation of Constrained Regression

Huda S. Ibrahim	Fadhaa M. Hashim
huda.sami2101m@coadec.uobaghdad.edu.iq	ali_fadhaa @coadec.uobaghdad.edu.iq
Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq	

Article Information

Article History:
Received: January, 9, 2024
Accepted: February, 28, 2024
Available Online: 15, May, 2025

Keywords:
Constrained regression, Hinge algorithm, inequality constraints, cone, Least Squares, conic projections, Quadratic programming.

Correspondence:
Huda S. Ibrahim
huda.sami2101m@coadec.uobaghdad.edu.iq

Abstract

Statisticians often use regression models like parametric, nonparametric, and semi-parametric models to represent economic and social phenomena. These models explain the relationships between different variables in these phenomena. One of the parametric model techniques is conic projection regression. It helps to find the most important slopes for multidimensional data using prior information about the regression's parameters to estimate the most efficient estimator. R algorithms, written in the R language, simplify this complex method. These algorithms are based on quadratic programming, which makes the estimations more accurate.

doi: https://doi.org/10.55562/jrucs.v57i1.12

1. Introduction

When the statistician faced many econometric, social, health, etc., Phenomena, they used different statistical methods to identify, estimate analyze these phenomena to have the ability to understand their trends and behaviors. Most statisticians work to study the mean variables that affect any phenomenon or case study by putting a proper model (Al-Salihi, 1992) and by using information like economic theory, health information, or any prior information and considering it as a restricted constraint and transforming it into mathematical models and the estimation of these. Regression parameters are also restricted in the border of the restricted constraint (Al-Naqash,1997). These constraints are variables, not constants, where the estimation of the model will lead to problems in estimation and inference (Al-Hasnawi and Al-Qaisi, 2002). To solve this problem, Gavin and Scruggs (2020) developed the restricted and unrestricted models and suggested a new algorithm to estimate the parameters by Lagrange Multiplier, and Mayer (2023) proposed using the quadratic algorithm in conic projections to estimate the parameters of the conic regression. In this research, we covered the three methods of estimation (OLS, Qp, and conic projection) by using simulation and real applicability.

2. Methodology Description

The process of estimating unknown parameters in a constrained and unconstrained regression model is showed in the below algorithm

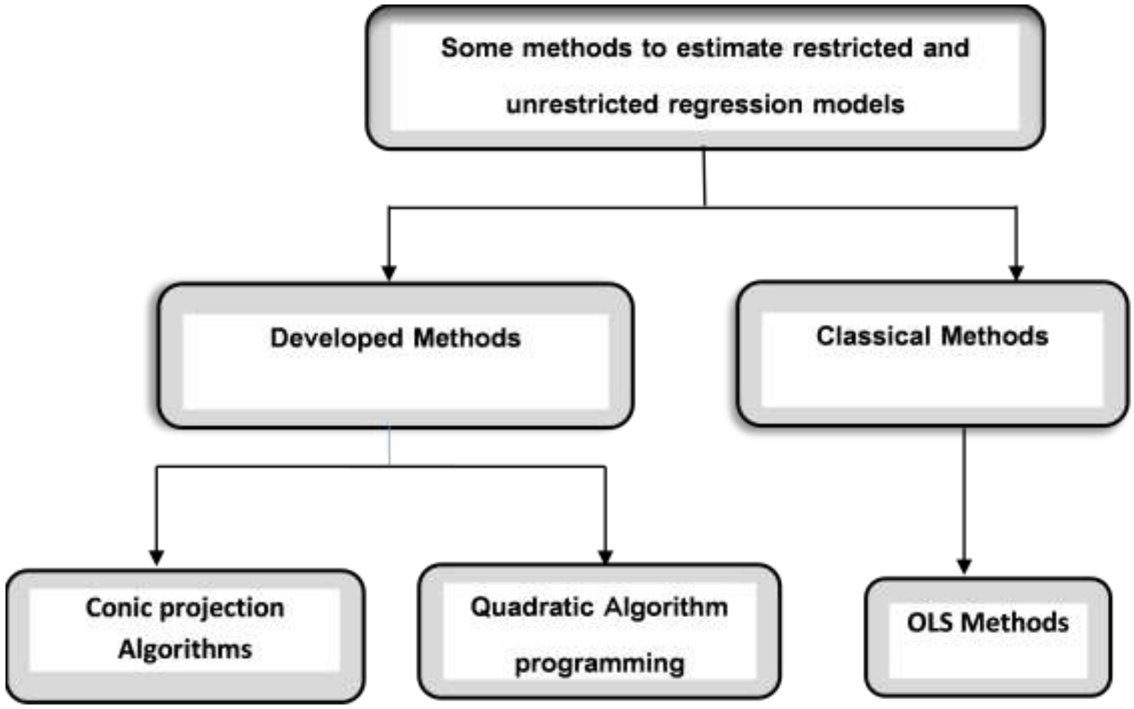


Chart No .(1): Shows the methods for estimating the parameters used in the research

And will be explained using these algorithms in the R language for statistical programming.

2.1. OLS Methods

One of the classical estimation methods for regression models is Ordinary Least Squares (OLS) for the model . (Mohamed,2014)(Gavin,2022)

$$Y = X\beta + U \quad (1)$$

$$\hat{\beta} = (x^T x)^{-1} x^T y \quad (2)$$

And the variance and var − cov (b_{ols}) are respectively

$$var - cov = S_e^2 (X^T X)^{-1} \quad (3)$$

$$S_e^2 = \frac{y^T y - B_{LS}^T X^T y}{n - k - 1} \quad (4)$$

2.2. The constraints

It is a mathematical relationship in the form of a linear equation between variables and specific resources, with two kinds of equal and unequal constraints as below:

$$(B_k)_{RLS} = (B_k)_{OLS} \quad if\ a \le (B_k)_{OLS} \le b \quad (5)$$

$$(B_k)_{RLS} = a \quad if\ (b_k)_{OLS} < a \quad (6)$$

$$= b\ if\ (B_k)_{OLS} > b \quad (7)$$

a: represents the minimum restriction
b: represents the upper limit of the restriction

(Al_Hashawi and Al_Qaisi,2002)

2.3.Estimation by Quadratic Programming (QP)

The QP is one of the mathematics programming that is used to determine a group of variables in an objective function to minimize or maximize when we mix previous information and the sample data to estimate the regression, parameters. The different restrictions about the parameters are represented as constraints, we assume the regression model is a QP problem. In R programming, the algorithm of QP estimation is

Table (1): Algorithm of QP estimation

• Inter the explanatory and response variables $(x_1, x_2, \dots \dots x_n), (y)$
• Determine the sample data (n)
• Generate a matrix of arranged constraints for the sample data as $RB \geq c$
• Calculate the constraint parameters $B^T AB - 2s^T B$

Source:(Liao and Meyer,2014)

2.4.Conic projection Estimation

Conic projection is a special case of QP that concerns minimizing objective function under different restrictions when we use conic projection - (Polar Conic) in many other procedures with transformed QP equations to conic projection problems and restricted regression parameter estimation by restriction matrix R, which contains a row of conical edges. These conical edges will be used in conic A and B side by side with a model and weight oriented where the researcher provides them, the restriction conic and conical edges matrices, and we will have a Q matrix of size $n * n$ by:

$$Q = X^T X \quad (8)$$

$$\beta = X(B - B_0) \quad (9)$$

$$y = (X^{-1})^T (s - QB_0) \quad (10)$$

Under the restriction

$$\min\|y - X\beta\| \quad (11)$$

$$R\beta \geq 0$$

Then the final step of estimation is

$$\| XB^{\sim} - XB \|^2 = (y - XB^{\wedge})^T (XB^{\wedge} - XB) - (y - XB^{\wedge})^T (XB^{\wedge} - XB)$$

(Meyer,2009 ,2013B)

2.5. Projection Matrix (Hat Matrix)

Hoaglin and Welsch(1977) studied In Linear regression model estimation hat matrix can be used to estimate the response variable vector with no need to estimate the regression model parameters and generate the estimation error vector by using the original data, as below:

$$e = y - y^{\wedge}$$

$$\hat{y} = XB^{\wedge}$$

$$\hat{\beta} = X(X^T X)^{-1} X^T y$$

Where $X(X^T X)^{-1} X^T$ is known as a projection matrix (hat matrix) for estimation

$$\therefore e = y - X(X^T X)^{-1} X^T y$$

$$e = (I - X(X^T X)^{-1} X^T) y$$

$$e = py$$

And p is known as a projection matrix (hat matrix) for error

In R programming the conic projection estimation algorithm can be written below

Table (2): Conic projection estimation algorithm

• Inter the explanatory and response variables $(x_1, x_2, \dots \dots x_n), (y)$
• Determine the sample data (n)
• Generate the Irreducible Restriction Matrix must provide for sample data and arrange according to $RB \geq c$
• Calculate the value of Polar conic by the function $p = \sum_{i=1}^{M} a^i g^i$
• Estimate weighted and un weighted restricted parameters and campier between them to select the best estimation according to Mean Square Error (MSE) for the model

Source: (Meyer,2009) (Meyer and Liao,2014)

3. The Simulation

By using simulation methods, we have built an experiment to test three different sizes of samples (50, 100, and 150) of data, the default values generated for the sample data between (0, 1), in a virtual form similar to real reality, and through the process of estimating the parameters mentioned above and making comparisons between the results according to the lowest MSE to reach for the best method of estimation.

We make a simple algorithm

Table (3): Simulation algorithm

• Choosing three data samples 50,100 and 150
• Generate five explanatory variables x_1, x_2, x_3, x_4, x_5 where $X_1 \sim Uniform(0,1)$, $X_2 = 0.6X_1 + e_1$, $X_3 = 0.6X_2 + e_2$, $X_4 \sim Bin(1,0.5)$, $X_5 \sim Bin(1,0.3)$
$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \beta_7 X_1 X_2 X_3 + \beta_8 X_4 + \beta_9 X_5 + \epsilon$ Where $\epsilon \sim N(0,0.2)$
Applied three methods of estimations OLS, QP, and CP

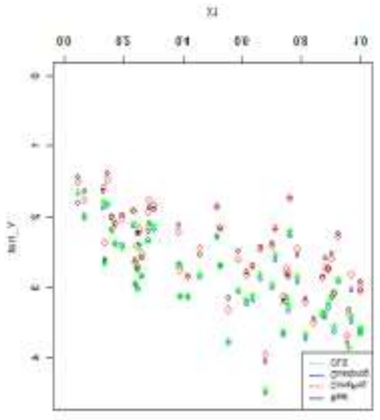
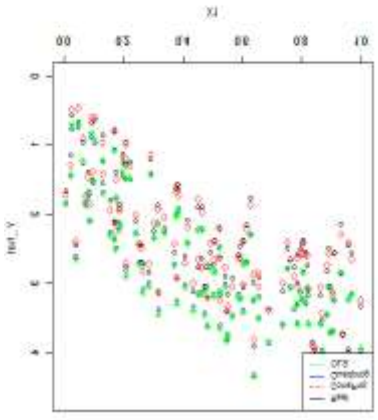
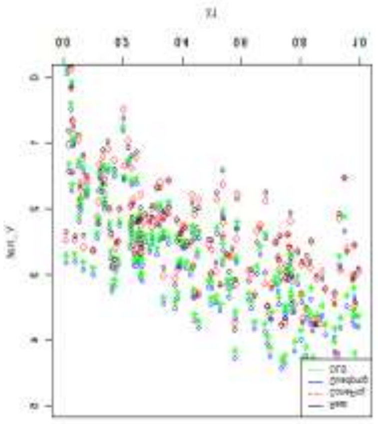
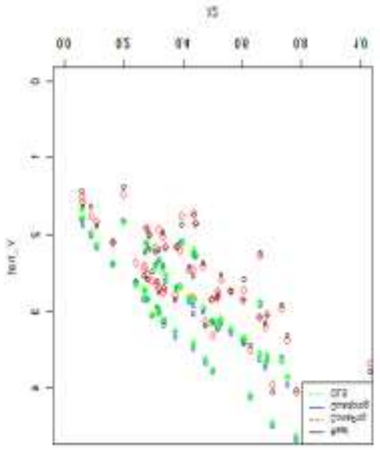
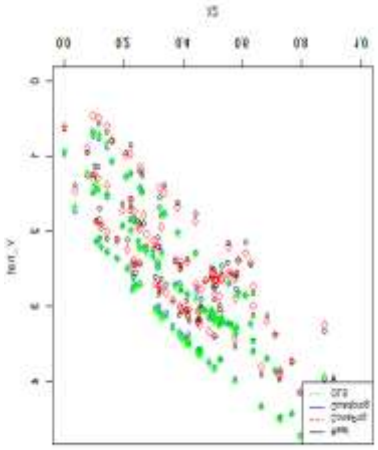
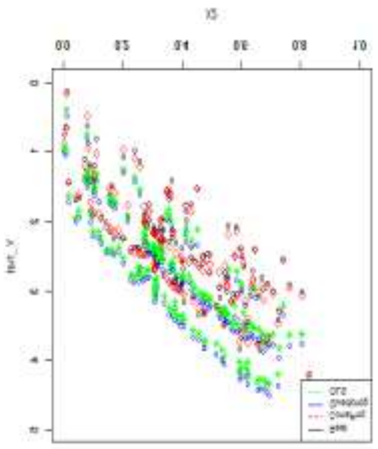
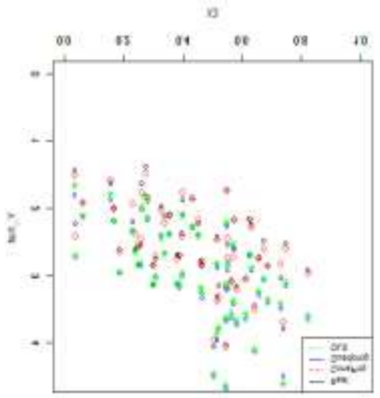
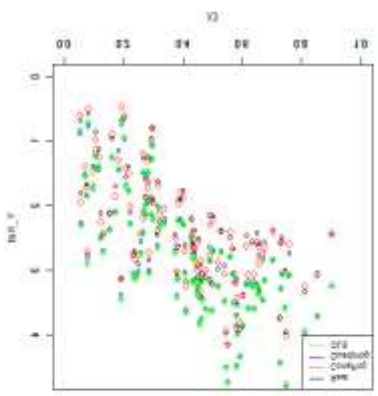
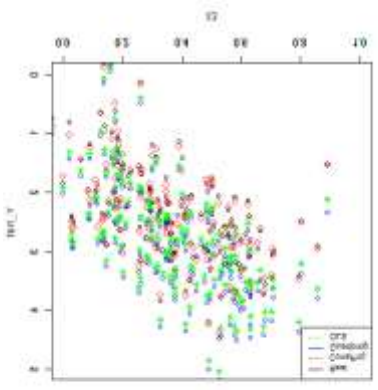
The results are :

Table (4): Parameters estimation by simulation

The samples	N=50			N=100			N=150		
Parameters	CP	QP	OLS	CP	QP	OLS	CP	QP	OLS
β_0	0.000824	0.002193	0.001875	0.000306	0.0016779	0.0013293	0.0001949	0.0013237	0.0011961
β_1	0.006662	0.002019	0.030951	0.00263722	0.00103117	0.00552975	0.00193287	0.00074028	0.00606446
β_2	0.011469	0.002018	0.025497	0.0095146	0.00110704	0.01640175	0.00597282	0.00077756	0.01120859
β_3	0.009822	0.002014	0.022879	0.00346609	0.0011005	0.00649434	0.00137277	0.00072965	0.00381172
β_4	0.003806	0.002003	0.056573	0.01029443	0.00100725	0.04340501	0.00726229	0.00068872	0.01634731
β_5	0.034562	0.002003	0.14403	0.013382	0.001006	0.033426	0.006134	0.000678	0.012582
β_6	0.128933	0.002002	0.142912	0.0638255	0.0010186	0.0772942	0.0307326	0.0006813	0.0744371
β_7	0.067983	0.002001	0.378793	0.02494235	0.00100147	0.09323367	0.03308175	0.00067012	0.07633881
β_8	0.000358	0.002	0.002408	9.92E-05	0.001	0.00096985	2.49E-05	7.79E-04	7.27E-04
β_9	0.000255	0.002	0.002451	6.63E-05	1.00E-03	1.05E-03	2.42E-05	6.68E-04	6.05E-04
MSE	0.008926	0.194695	0.188873	0.00838924	0.21263558	0.18318001	0.00948352	0.22970274	0.18777656

From Table (4) we can see easily that MSE for all CP estimators is less than the other estimators, i.e. the CP estimation method is more efficient than OLS and QP.

Table (5): Represent MSE of parameters values by using quadratic programming and Conic projection with n= 50, 100, 150

n \ v	50	100	150
X1			
X2			
X3			

4. Applicable part

Using the results in the simulation part can be able to apply it to real case studies. The researchers have data about the level of forced expiratory volume (FEV) taken from a pulmonary function unit / private nursing home hospital for (x_5)male and female (x_4) smokers and nonsmokers with ages between 7 and 69 years (x_1) with there weights (x_3) and heights(x_2), by using facts, scientific research, and previous studies on Iraqi health field, we can build a restriction on the data with response variables (y) that represent the FEV.

The results of estimation by CP is in the table below:

Table (6): Real data estimation by CP

Regression Parameters Estimation	
β_0	6.280402e-01
β_1	1.998401e-14
β_2	3.675577
β_3	7.105427e-15
β_4	-8.881784e-16
β_5	-1.620926e-14
β_6	2.699537e-1
β_7	-2.960189
β_8	3.139213e-01
β_9	-9.642468e-01
$y = 0.628 + 1.998\text{e} - 14x_1 + 3.676x_2 + 7.105e - 15x_3 - 8.88e - 16x_1x_2 - 1.621e - 14x_1x_3 + 0.2699x_2x_3 - 2.96x_1x_2x_3 + 0.3139x_4 - 0.964x_5$	

From table (6) we can see:

- (β_1 , β_3, β_4, β_5) are very close to Zero
- (β_2, β_6, β_8) are positive values that lead to direct effects.
- (β_7, β_9) are negative values that lead to reverse effects.

5. Conclusions

- The CP estimators are the best estimator than OLS and QP
- $\because$ (β_1 , β_3) ≈ 0 Then x_1(Age)and x_3(weight) will be excluded from the model.
- $\because$ (β_4 , β_5) ≈ 0 Then the contrasts x_1x_2 and x_1x_3 will be excluded from the model.
- x_2, x_4 Have positive effects on the model or i.e. on the FEV.
- The contrast x_2x_3 has a positive effect on the model or i.e. on the FEV.
- The contrast ($x_1x_2x_3$) has negative effects on the model or i.e. on the FEV.
- x_5 Has a negative effect on the model or i.e. on the FEV.

6. Recommendations

- Expanding the study of the conical projections estimation from unequal constraints to equal constraints on the parameters of the linear regression model.
- Expanding this method to estimate nonparametric and semiparametric regression models under equal and unequal constraints of the models.

References

[1] Al –Salhi, Z(1992),” Identical and non identical Constraints in regression analysis an applied study of the consumption function in Iraq”, College of Administration and Economics , University of Baghdad.

[2] Al-Hashawi , A & Al-Qaisi , B (2002), “ Advanced Economic Econometrics Theory and Practice”, National Library, House of Books and Documents, Baghdad 552, (2002).

[3] Al-Napash ,I(1997),”Planning Residential Neighborhood Electricity Distribution Networks” ,college of Administration and Economics University of Baghdad.

[4] Gavin Henri P, (2022) .” Constrained Linear Least Squares” Duke University , Spring.

[5] Gavin Henri P, Scruggs Jeffrey T , (2020) . “Constrained Optimization Using Lagrange Multipliers”, Spring 2020

[6] Hoaglin David C & Welsch Roy E(1978),"The Hat Matrix in regression and ANOVA", The American Statistician, Vol. 32, No. 1 (Feb., 1978), pp. 17-22.

[7] Meyer C Mary (2013b)." A simple New Algorithm for Quadratic programming with Applications in statistics" communications in statistics- simulation and computation, 42(5).

[8] Meyer C Mary, (2009). " An Algorithm for Quadratic Programming with Applications in Statistics"

[9] Meyer C Mary, Liao Xiyue (2014)." Coneproj :An R Package for the Primal or Dual Cone Projection with Routines for Constrained Regression", Journal of statistical software, Vol. 61, No. 12.

[10] Meyer C Mary, Liao Xiyue (2023). "Primal or Dual Cone Projection with Routines for Constrained Regression" Date 2023-10-13 ,Version 1.17

[11] Muhammad, J., (2014), "Using lest squares and restricted least squares in estimating the first - order auto regressive parameter AR(1) (Simulation Study).", College of Administration and Economics , University of Baghdad Department of statistics.

مجلة كلية الرافدين الجامعة للعلوم (2025)؛ العدد 57؛ 136- 143

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

استخدام الاسقاطات المخروطية والبرمجة التربيعية في تقدير الانحدار المقيد

هدى سامي ابراهيم	م. د. فضاء مزهر هاشم
huda.sami2101m@coadec.uobaghdad.edu.iq	ali_fadhaa @coadec.uobaghdad.edu.iq

قسم الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد، بغداد، العراق

المستخلص

يستخدم الاحصائيون عادة نماذج الانحدار من قبيل المعلمية واللامعلمية والشبه المعلمية لتحليل وتوضيح الظواهر الاجتماعية والاقتصادية المختلفة , وهذه النماذج عموما تفسر العلاقة بين مختلف المتغيرات الموجودة في هذه الظواهر . يعتبر انموذج الانحدار المخروطي هو احد تقنيات النماذج المعلمية الخطية , حيث يساعد في ايجاد اهم الاسقاطات لمختلف الابعاد التي تصاحب البيانات باستخدام المعلومات الاولية او المسبقة حول معالم الانحدار الخطي من اجل تقدير اكثر المعالم كفاءة ودقة , ولكنه يعتبر من الطرق المعقدة , وباستخدام برنامج R وخوارزمياته المختلفة فان عملية التقدير بهذا الاسلوب قد اصبحت اسهل بكثير حيث ستعتمد هذه الخوارزميات بالأصل على البرمجة التربيعية والتي ستساعد على تقدير المعالم بصورة كفوءة .

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث:9/1/2024
تاريخ قبول البحث:28/2/2024
تاريخ رفع البحث على الموقع: 15/5/2025

الكلمات المفتاحية:
الانحدار المقيد، الخوارزمية المفصلة، قيود عدم المساواة، المخروط، المربعات الصغرى الاعتيادية، الاسقاطات المخروطية، البرمجة التربيعية

للمراسلة:
هدى سامي ابراهيم
huda.sami2101m@coadec.uobaghdad.edu.iq

doi: https://doi.org/10.55562/jrucs.v57i1.12