Almost Unbiased Liu Principal Component Estimator In The Presence Of Multicollinearity and Autocorrelation

Prof. Ahmed A.E.
Dep. of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University

Prof. Hassan M.Ali
Dep. of Statistics, Faculty of Commerce, Zagazig University

Amal, H. A.
MSc. Student, Dep. of Statistics, Faculty of Commerce, Zagazig University

Abstract. In this article, a new class of estimator called the Almost Unbiased Liu Principal Component Estimator (AULPCR) for the multiple linear regression model with autocorrelated errors in the presence of multicollinearity problem will be suggested. The properties of the proposed estimator is discussed.

Keywords: Principal components – Liu estimator – Autocorrelation – Multicollinearity – Unbiasedness – Multiple Linear regression model.

1. Introduction

Consider the following multiple linear regression model

$$Y = X\beta + \varepsilon \tag{1}$$

Where Y is an $n \times 1$ vector of observations, X is an $n \times p$ design matrix with rank p, β is a $p \times 1$ vector of regression coefficient and ε is an $n \times 1$ vector of errors with expectation $E(\varepsilon) = 0$ and variance covariance matrix $Var(\varepsilon) = \sigma^2\Omega$, Ω is a known $n \times n$ positive definite matrix .

Since, Ω is a symmetrical, positive definite matrix, there exists an orthogonal matrix P such that $P'P = \Omega$. On premultiplying model (1) by P^{-1}, we have

$$P^{-1}Y = P^{-1}X\beta + P^{-1}\varepsilon$$

Let, $Y^* = P^{-1}Y$, $X^* = P^{-1}X$ and $\varepsilon^* = P^{-1}\varepsilon$ then $E\left(\varepsilon^*\right) = 0$ and

$Var\left(\varepsilon^*\right) = \sigma^2 I_n$

Therefore the transformed model can be written as:

$$Y^* = X^*\beta + \varepsilon^* \tag{2}$$

The OLS estimator for the model (2) is

$$\tilde{B}_{GLS} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

That is known by the generalized least squares (GLS) estimator of β. The GLS estimator is the best linear unbiased estimator of β. This estimator performs very well until the explanatory variables show multicollinearity, because of the large total variance. (Alheety and Kibria, 2009).

Several authors have studied the problem of multicollinearity and autocorrelation simultaneously. Alheety and Kibria, (2009) proposed the almost unbiased liu (AUL) estimator of the linear regression model as;

$$\tilde{\beta}_{AUL} = \left[I + \left(X'\Omega^{-1}X + I\right)^{-1}(1-d)\right]\tilde{\beta}_{LE}$$

$$\text{Or } \tilde{\beta}_{AUL} = \left[I - \left(X'\Omega^{-1}X + I\right)^{-2}(1-d)^2\right]\tilde{\beta}_{GLS}$$

Where, $\tilde{\beta}_{LE} = (X'\Omega^{-1}X + I)^{-1}(X'\Omega^{-1}X + dI)\tilde{\beta}_{GLS}$, $\tilde{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}Y$

Siray, et al, (2014) combined the approaches of the $r-k$ class estimator proposed by Baye and Parker (1984) and the GLS estimator to get the estimator:

$$\tilde{\beta}_r(k) = T_r(T_r'X'\Omega^{-1}XT_r + kI_r)^{-1}T_r'X'\Omega^{-1}Y \quad k > 0 \qquad (3)$$

Where, $T_r = [t_1, t_2, t_3, \ldots\ldots, t_r]$ be $p \times r$ orthogonal matrix after deleting last $p-r$ columns from T matrix, where $r \le p$ and $T = [t_1, t_2, t_3, \ldots\ldots, t_p]$ be orthogonal matrix with $T'X'\Omega^{-1}XT = \Delta$ where $\Delta = diag(\lambda_1, \lambda_2, \ldots\ldots, \lambda_p)$ is the diagonal matrix of the eigenvalues of $X'\Omega^{-1}X$.

Huang and Yang (2015) generalized the principal component two parameter (PCTP) estimator that proposed by Chang and Yang (2012) for dealing with multicollinearity and autocorrelation problems simultaneously which given by;

$$\tilde{B}_r(k,d) = T_r(T_r'X'\Omega^{-1}XT_r + I_r)^{-1}(T_r'X'\Omega^{-1}XT_r + dI_r)(T_r'X'\Omega^{-1}XT_r + kI_r)^{-1}T_r'X'\Omega^{-1}Y \qquad (4)$$

Where, $k > 0$, $0 < d < 1$ and T_r is defined as in equation (3)

The primary aim of this article is to introduce a new class of estimator where GLSE, PCRE and AULE will be derived as special cases. The article is organized as follows. In section 2, the new estimator is introduced. In Section 3, some properties of the new estimator is discussed. In section 4, application study is performed.

2. The Proposed Estimator

Method (1):

The transformed model in (2) can be written as

$$Y^{*} = X^{*}T_rT_r'\beta + X^{*}T_{p-r}T_{p-r}'\beta + \varepsilon^{*} \qquad (5)$$

The idea of PCR corresponds to the transition from the model (5) to the reduced model

$$Y^{*} = X^{*}T_rT_r'\beta + \varepsilon^{*} \ , \qquad (6)$$

By applying the steps of AUL estimator that proposed by Alheety and Kibria, (2009) to the reduced model (6) as follows:

Step 1:

$$\varepsilon^{*\prime}\varepsilon^{*} = \left(Y^{*} - X^{*}T_rT_r'\hat{\beta}\right)'\left(Y - X^{*}T_rT_r'\hat{\beta}\right) + \left(T_r'\hat{\beta} - d\tilde{\beta}^{*}\right)'\left(T_r'\hat{\beta} - d\tilde{\beta}^{*}\right)$$

Where $\tilde{\beta}^{*} = (\mathrm{T}_r' X^{*\prime}X^{*}\mathrm{T}_r)^{-1}\mathrm{T}_r' X^{*\prime}Y^{*}$

For simplicity, let, $X^{*}T_r = X^{\bullet}$ and $T_r'\hat{\beta} = \hat{\beta}^{\bullet}$ then, $\varepsilon^{*\prime}\varepsilon^{*}$ can be written as

$$\varepsilon^{*\prime}\varepsilon^{*} = \left(Y^{*} - X^{\bullet}\hat{\beta}^{\bullet}\right)'\left(Y - X^{\bullet}\hat{\beta}^{\bullet}\right) + \left(\hat{\beta}^{\bullet} - d\tilde{\beta}^{*}\right)'\left(\hat{\beta}^{\bullet} - d\tilde{\beta}^{*}\right)$$

$$\frac{\partial\varepsilon^{*\prime}\varepsilon^{*}}{\partial\hat{\beta}^{\bullet}} = 2\hat{\beta}^{\bullet}X^{\bullet\prime}X^{\bullet} - 2X^{\bullet\prime}Y^{*} + 2\hat{\beta}^{\bullet} - 2d\tilde{\beta}^{*}$$

$$\frac{\partial\varepsilon^{*\prime}\varepsilon^{*}}{\partial\hat{\beta}^{\bullet}} = 2\hat{\beta}^{\bullet}X^{\bullet\prime}X^{\bullet} - 2X^{\bullet\prime}Y^{*} + 2\hat{\beta}^{\bullet} - 2d\tilde{\beta}^{*} = 0$$

$$\hat{\beta}^{\bullet}(X^{\bullet\prime}X^{\bullet} + I) = X^{\bullet\prime}Y^{*} + d\tilde{\beta}^{*}$$

$$\hat{\beta}^{*}=(X^{*\prime}X^{*}+I)^{-1}(X^{*\prime}Y^{*}+d\tilde{\beta}^{*})$$

By substituting $X^{*}=X^{*}T_r$, $\hat{\beta}^{*}=T_r'\hat{\beta}$ and $\tilde{\beta}^{*}=(T_r'X^{*\prime}X^{*}T_r)^{-1}T_r'X^{*\prime}Y^{*}$, then

$$T'\hat{\beta}=(T_r'X'^{*}X^{*}T_r+I_r)^{-1}(T_r'X'^{*}Y^{*}+d(T_r'X'^{*}X^{*}T_r)^{-1}T_r'X'^{*}Y^{*})$$

$$T'\hat{\beta}=(T_r'X^{*\prime}X^{*}T_r+I_r)^{-1}(T_r'X^{*\prime}X^{*}T_r+dI_r)(T_r'X^{*\prime}X^{*}T_r)^{-1}T_r'X^{*\prime}Y^{*}$$

Since, $Y^{*}=P^{-1}Y$, $X^{*}=P^{-1}X$ and $P'P=\Omega$, then

$$\begin{aligned}\tilde{\beta}_{LEPCR}&=T_r(T_r'X'\Omega^{-1}XT_r+I_r)^{-1}(T_r'X'\Omega^{-1}XT_r+dI_r)(T_r'X'\Omega^{-1}XT_r)^{-1}T_r'X'\Omega^{-1}Y\\&=T_r(T_r'X'\Omega^{-1}XT_r+I_r)^{-1}(T_r'X'\Omega^{-1}XT_r+dI_r)T_r'\tilde{\beta}_{PCR}\end{aligned}\qquad(7)$$

Where, $\tilde{\beta}_{PCR}=T_r(T_r'X'\Omega^{-1}XT_r)^{-1}T_r'X'\Omega^{-1}Y$

Step 2:

$\tilde{\beta}_{LEPCR}$ in equation (7) can be written as:

$$\begin{aligned}\tilde{\beta}_{LEPCR}&=T_rT_r'\tilde{\beta}_{PCR}-T_r(1-d)(T_r'X'\Omega^{-1}XT_r+I_r)^{-1}T_r'\tilde{\beta}_{PCR}\\&=\left[T_rT_r'-T_r(1-d)(T_r'X'\Omega^{-1}XT_r+I)^{-1}T_r'\right]\tilde{\beta}_{PCR}\end{aligned}$$

(8)

Step 3:

Since, $bias\ (\tilde{\beta}_{LEPCR})=-\left[T_r(1-d)(T_r'X'\Omega^{-1}XT_r+I_r)^{-1}T_r'+T_{p-r}T_{p-r}'\right]\beta$, then following Kadiyala (1984), the bias of corrected LEPCR estimator of β is given by

$$\tilde{\beta}_{LEPCR}^{*}=\tilde{\beta}_{LEPCR}+\left[T_r(1-d)(T_r'X'\Omega^{-1}XT_r+I_r)^{-1}T_r'+T_{p-r}T_{p-r}'\right]\beta \qquad (9)$$

Step 4:

Following Ohtani (1986) by replacing the β in (9) by the biased estimator $\tilde{\beta}_{LEPCR}$ in (8), we have

$$\tilde{\beta}_{AULPCR} = \tilde{\beta}_{LEPCR} + \left[T_r(1-d)(T_r'X'\Omega^{-1}XT_r + I_r)^{-1}T_r' + T_{p-r}T_{p-r}'\right]\tilde{\beta}_{LEPCR}$$

$$= \left[I + T_r(1-d)(T_r'X'\Omega^{-1}XT_r + I_r)^{-1}T_r'\right]\tilde{\beta}_{LEPCR}$$

$$= \left[I + T_r(1-d)(T_r'X'\Omega^{-1}XT_r + I_r)^{-1}T_r'\right]\left[T_rT_r' - T_r(1-d)(T_r'X'\Omega^{-1}XT_r + I_r)^{-1}T_r'\right]\tilde{\beta}_{PCR}$$

$$= \left[T_rT_r' - T_r(1-d)^2(T_r'X'\Omega^{-1}XT_r + I_r)^{-2}T_r'\right]\tilde{\beta}_{PCR}$$

Then the AULPCR estimator is given by

$$\tilde{\beta}_{AULPCR}(r,d) = \left[T_rT_r' - T_r(1-d)^2\left(T_r'X'\Omega^{-1}XT_r + I_r\right)^{-2}T_r'\right]\tilde{\beta}_{PCR}$$

Where $\tilde{\beta}_{PCR} = T_r(T_r'X'\Omega^{-1}XT_r)^{-1}T_r'X'\Omega^{-1}Y$

Also, the proposed estimator can be obtained by the following method.

Method (2):

Following Siray, et al (2014), the $r-k$ class estimator $\tilde{\beta}_r(k)$ in equation (3) can be written as

$$\tilde{\beta}_r(k) = T_rT_r'\tilde{\beta}_{RR}$$

Where $\tilde{\beta}_{RR} = (X'\Omega^{-1}X + kI)^{-1}X'\Omega^{-1}Y$ and also can be written as:

$$\tilde{\beta}_{RR} = T(T'X'\Omega^{-1}XT + kI)^{-1}T'X'\Omega^{-1}Y$$

Since, $T_r T_r'T(X'\Omega^{-1}X)^{-1}T' = T_r(T_r'X'\Omega^{-1}XT_r)^{-1}T_r'$

Then, $\tilde{\beta}_r(k) = T_rT_r'\tilde{\beta}_{RR}$

$$=T_r T_r' T\,(T'X'\Omega^{-1}XT+kI)^{-1}T'X'\Omega^{-1}Y$$

$$=T_r\,(T_r'X'\Omega^{-1}XT_r+kI_r)^{-1}T_r'X'\Omega^{-1}Y$$

Following Huang and Yang (2015), the generalized (PCTP) estimator in equation (4) can be written as

$$\tilde{\beta}_r(k,d)=T_rT_r'\tilde{\beta}(k,d)$$

Where, $\tilde{\beta}(k,d)=(X'\Omega^{-1}X+I)^{-1}(X'\Omega^{-1}X+dI)(X'\Omega^{-1}X+kI)^{-1}X'\Omega^{-1}Y$

which can be rewritten as:

$$\tilde{\beta}(k,d)=T\,(T'X'\Omega^{-1}XT+I)^{-1}(T'X'\Omega^{-1}XT+dI)(T'X'\Omega^{-1}XT+kI)^{-1}T'X'\Omega^{-1}Y$$

Since, $T_r\,T_r'T(X'\Omega^{-1}X)^{-1}T'=T_r\,(T_r'X'\Omega^{-1}XT_r)^{-1}T_r'$

Then, $\tilde{\beta}_r(k,d)=T_rT_r'\tilde{\beta}(k,d)$

$$=T_r(T_r'X'\Omega^{-1}XT_r+I_r)^{-1}(T_r'X'\Omega^{-1}XT_r+dI_r)(T_r'X'\Omega^{-1}XT_r+kI_r)^{-1}T_r'X'\Omega^{-1}Y$$

Hence, following the above procedure and the definition of almost unbiased liu (AUL) estimator proposed by Alheety and Kibria, (2009), the proposed estimator will be obtained as;

$$\tilde{\beta}_{AULPCR}(r,d)=T_rT_r'\tilde{\beta}_{AUL}$$

$$=T_rT_r'\left[I-(1-d)^2\left(X'\Omega^{-1}X+I\right)^{-2}\right]\tilde{\beta}_{GLS} \qquad (10)$$

$$=T_rT_r'\left[I-T\,(1-d)^2\left(T'X'\Omega^{-1}XT+I\right)^{-2}T'\right]T\left(T'X'\Omega^{-1}XT\right)^{-1}T'X'\Omega^{-1}Y$$

$$=\left[T_rT_r'-T_r\,(1-d)^2\left(T_r'X'\Omega^{-1}XT_r+I_r\right)^{-2}T_r'\right]\tilde{\beta}_{PCR}$$

So, the Almost Unbiased Liu Principal Component Estimator $\tilde{\beta}_{AULPCR}(r,d)$ can be written as:

$$\tilde{\beta}_{AULPCR}(r,d)=T_rT_r'\left[I-(1-d)^2\left(X'\Omega^{-1}X+I\right)^{-2}\right]\tilde{\beta}_{GLS}$$

Or $\tilde{\beta}_{AULPCR}(r,d)=\left[T_rT_r'-T_r(1-d)^2\left(T_r'X'\Omega^{-1}XT_r+I_r\right)^{-2}T_r'\right]\tilde{\beta}_{PCR}$

The AULPCR estimator $\tilde{\beta}_{AULPCR}(r,d)$ is a general estimator which includes the GLS estimator, the PCR estimator and the AUL estimator as special cases:

$\tilde{B}_p(p,1)=\tilde{B}_{GLS}=(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$, $(r=p$ and $d=1)$.

$\tilde{B}_r(r,1)=\tilde{\beta}_{PCR}=T_r(T_r'X'\Omega^{-1}XT_r)^{-1}T_r'X'\Omega^{-1}Y$, $(d=1)$.

$\tilde{\beta}_p(p,d)=\tilde{\beta}_{AUL}=\left[I-\left(X'\Omega^{-1}X+I\right)^{-2}(1-d)^2\right]\tilde{\beta}_{GLS}$, $(r=p)$.

3. Properties of the proposed estimator

1- The Bias of AULPCR estimator.

Since, $\tilde{\beta}_{AULPCR}(r,d)=T_rT_r'\left[I-(1-d)^2\left(X'\Omega^{-1}X+I\right)^{-2}\right]\tilde{\beta}_{GLS}$

Let $\Gamma=X'\Omega^{-1}X$ and $H_d=I-(1-d)^2(\Gamma+I)^{-2}$

Then, $\tilde{\beta}_{AULPCR}=T_rT_r'H_d\tilde{\beta}_{GLS}$

Now since, $Bias(\tilde{\beta}_{AULPCR}(r,d))=E\left(\tilde{\beta}_{AULPCR}(r,d)\right)-\beta$

Then, $E\left(\tilde{\beta}_{AULPCR}\right)=E\left(T_rT_r'H_d\tilde{\beta}_{GLS}\right)=T_rT_r'H_d\beta$

$$\therefore Bias(\tilde{\beta}_{AULPCR}(r,d)) = (T_r T_r' H_d - I)\beta \tag{11}$$

2- The Variance-Covariance Matrix of AULPCR estimator.

Since, $\tilde{\beta}_{AULPCR} = T_r T_r' H_d \tilde{\beta}_{GLS}$

So, $Var(\tilde{\beta}_{AULPCR}(r,d)) = Var(T_r T_r' H_d \tilde{\beta}_{GLS}) = T_r T_r' H_d Var(\tilde{\beta}_{GLS}) H_d' T_r T_r'$

Since $Var(\tilde{\beta}_{GLS}) = \sigma^2 \Gamma^{-1}$

Then, $Var(\tilde{\beta}_{AULPCR}(r,d)) = \sigma^2 (A\Gamma^{-1}A')$ (12)

Where, $A = T_r T_r' H_d$

3- Mean Square Error of AULPCR estimator.

Since,

$$MSE(\tilde{\beta}_{AULPCR}(r,d)) = Var(\tilde{\beta}_{AULPCR}(r,d)) + \left[Bias(\tilde{\beta}_{AULPCR}(r,d))\right]\left[Bias(\tilde{\beta}_{AULPCR}(r,d))\right]'$$

Then, from equations (11&12), it can be concluded that,

$$MSE(\tilde{\beta}_{AULPCR}(r,d)) = \sigma^2 A\Gamma^{-1}A' + V\beta\beta'V'$$

Where, $V = (T_r T_r' H_d - I)$

4- Application Study.

In the presence of multicollinearity among explanatory variables, OLSE becomes unstable and shows undesirable properties. It is quite common in applied work to have autocorrelaion in error terms which reduce the efficiency of the OLSE. It has been observed that the problem of autocorrelation and multicollinearity arise simultaneously in several cases.

In this section, the performance of AULPCR estimator is compared with the GLS estimator, the PCR estimator and the AUL estimator in MSE as criteria for comparison.

1- Real Data Application.

The data have been taken from the United Nations Conference on Trade and Development (UNCTAD[1]) which restricted to Egypt from 1980-2018.

- Response is: Foreign Direct Investment FDI (Y)
- Predictors are :

1- Balance of Payment (X_1)

2- Gross Domestic Product GDP(X_2)

3- Government Final Consumption Expenditure (X_3)

2- Data Analysis.

The data has been analyzed using SPSS (ver,22) and R programs. The following results are obtained:

Table (1): Correlation Matrix.

Correlations

		FDI	balance of payment	GDP	government expenditure
	Pearson Correlation	1	-.425-**	.640**	.618**
FDI	Sig. (2-tailed)		.007	.000	.000
	N	39	39	39	39

1 https://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx

balance of payment	Pearson Correlation		1	-.721-**	-.736-**
	Sig. (2-tailed)			.000	.000
	N			39	39
GDP	Pearson Correlation			1	.990**
	Sig. (2-tailed)				.000
	N				39
government expenditure	Pearson Correlation				1
	Sig. (2-tailed)				
	N				39

**. Correlation is significant at the 0.01 level (2-tailed).

Table (2): ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	176572287.150	3	58857429.050	8.598	.000[b]
	Residual	239605493.157	35	6845871.233		
	Total	416177780.308	38			

Table (3): Coefficients

Model		Unstandardized Coefficients		Sig.	Collinearity Statistics	
		B	Std. Error		Tolerance	VIF
1	(Constant)	423.131	727.117	.564		
	balance of payment	.027	.127	.834	.454	2.200
	GDP	.052	.033	.126	.019	51.871
	government expenditure	-.254-	.305	.411	.018	54.385

From table (1), it can be seen that there is a correlation between FDI and the predictors (Balance of Payment, GDP and Government Expenditure). Also, the problem of multicollinearity can be observed between (X_1, X_2) , between (X_1, X_3) and between (X_2, X_3).

Based on tables (2, 3), the effect of the problem of multicollinearity can be seen where the coefficients of the predictors are not significant.

The value of Durbin Watson (DW) statistic is found to be 0.642 which indicates that there is a positive autocorrelation at significance level 0.05 with two limits of the critical value being $d_L = 1.328$ and $d_U = 1.658$. Using the results of DW test for AR (1) serial correlation then, the error structure follows AR (1) process with estimated ρ to be 0.68. Thus the Ω matrix can be constructed using the formula $\Omega = \rho^{|i-j|}$.

- Applying the methods of estimation on the data.

The eigenvalues of $X'\Omega^{-1}X$ are 2.998, $1.006e^{-03}$, $6.036e^{-17}$. We choose r =1 and a value of k and d to be 2 and 4 respectively.

The result of MSE of the GLS estimator, the PCR estimator and the AUL estimator and AULPCR Estimator are obtained as follows:

Table (4): MSE of different methods of estimation.

Method	**MSE**
The GLS estimator	4.27158
The PCR estimator	1.77379
The AUL Estimator	3.421994
The AULPCR Estimator	1.77322

From Table (4) we find that the estimated MSE of the proposed method AULPCR is smaller than the other estimators. So, based on the real data application the AULPCR and the PCR estimator are performing equally well rather than the other estimators.

References

1- Alheety, M. I., & Kibria, B. G. (2009). On the Liu and almost unbiased Liu estimators in the presence of multicollinearity with heteroscedastic or correlated errors. *Surveys in Mathematics and its Applications*, *4*, 155-167.

2- Baye MR & Parker DF (1984).Combining ridge and principal component regression: a money demand illustration. Commun. Stat Theory Methods 13,197–205

3- Chang, X., & Yang, H. (2012). Combining two-parameter and principal component regression estimators. *Statistical Papers*, *53*(3), 549-562.

4- Huang, J., & Yang, H. (2015). On a principal component two-parameter estimator in linear model with autocorrelated errors. *Statistical Papers*, *56*(1), 217-230.

5- Kadiyala, K. (1984). A class almost unbiased and efficient estimators of regression coefficients, Economics Letter, 16, 293-296.

6- Ohtani, K. (1986). On small sample properties of the almost unbiased generalized ridge estimator (1986), Communications in Statistics-Theory and Methods, 15, 1571-1578.

7- Şiray, G. Ü., Kaçıranlar, S., & Sakallıoğlu, S. (2014). r − k Class estimator in the linear regression model with correlated errors. *Statistical Papers*, *55*(2), 393-407.