Aitken’s Generalized Least Square Method for Estimating Parameter of Demand Function of Animal Protein In Indonesia

Fitria Virgantari¹, Hagni Wijayanti¹, Sonny Koeshendrajana²

¹Departement of Mathematics, Pakuan University
²Agency for Research and Manpower Resource of the Marine Affairs and Fisheries

Email: fitria.virgantari@unpak.ac.id

Abstract. Ordinary least square (OLS) method is the most popular and commonly technique used to estimate numerical values of parameters of selected regression models. This was due to its unbiasedness property. However, when elements of dependent variable have unequal variances and/or correlated each other, there is no guarantee that the OLS estimator will show the most efficient within the class of linear unbiased estimators. For conditions generally encountered, GLS method is proposed an estimation procedure which yields coefficient estimators at least asymptotically more efficient than single equation OLS estimator. This method is derived by Aitken and it’s named Aitken GLS. This paper reported a study of application of Aitken’s GLS method for estimating parameter of demand function of animal protein in Indonesia of which have a system equation. This system of equation causes violation of the assumptions of homoscedasticity and independency of estimated parameters. Secondary data obtained from the Central Bureau of Statistics 2016 in 34 provinces in Indonesia was used in this study. Animal food was grouped in terms of fish, meat, eggs and milk. Results showed that error of those three equations were correlated. It suggests that GLS method should be used. Normality and homogenity assumptions were not violated. The determination coefficient was 99%, indicating that the method was very good.

1. Introduction

The ordinary least square (OLS) method is the most popular and the oldest technique to estimate the numerical values of the parameters of regression model. It’s known that the classical conditions (normality, homoscedasticity, non-multicollinearity) need not hold in practice. Although these conditions have no effect on the OLS method, they do affect the properties of the OLS estimators and resulting test statistics. In particular, when the elements of the dependence variable have unequal variances and/or are correlated, the variance is no longer a scalar variance-covariance matrix, and hence there is no guarantee that the OLS estimator is the most efficient within the class of linear unbiased (or the class of unbiased) estimators.

In fact, there are many situations where there are several correlated error in several equations error so that the assumptions of the OLS method not to be held. This situations often found in the food consumption/demand study, which consists of two or more equations as a system of equations that are related to one another in some particular way [1]. It is only under special conditions, stated explicitly, that classical OLS applied equation-by-equation yields efficient coefficient estimators. For conditions generally encountered, Generalized least square (GLS) method is proposed an estimation procedure
which yields coefficient estimators at least asymptotically more efficient than single equation OLS estimators [2]. This method is derived by Aitken and it is named Aitken GLS. This paper presents a study of application of Aitken’s GLS method for estimating parameter of demand function of animal protein in Indonesia which have a system equation, so it may cause assumptions of homoscedasticity and independence are violated.

2. Ordinary Least Square
In statistics, ordinary least squares (OLS) is a common method for estimating the unknown parameters in a linear regression model. This method minimizes the sum of squared vertical distances between the observed responses in the data set and the responses predicted by the linear approximation. The OLS estimator is consistent when the regressors are exogenous and there is no multicollinearity, and optimal in the class of linear unbiased estimators when the errors are homoscedastic and serially uncorrelated. Under these conditions, the method of OLS provides minimum-variance mean-unbiased estimation when the errors have finite variances.

The matrix form of linear regression model is:

\[ Y = X\beta + \epsilon \]  \hspace{1cm} (1)

with \( X \) is independent variable with size of \( nxk \), \( Y \) is vector of dependent variable with size of \( nx1 \), \( \beta \) is vector parameter with size of \( kx1 \), \( \epsilon \) is vector error with size of \( nx1 \); \( n \) is number of observation, and \( k=p+1 \) is number of parameter.

In regression model, it’s assumed that \( \epsilon \) is random variable which drawn from normal distribution with mean zero and variance \( \sigma^2 \); \( \epsilon_i \) and \( \epsilon_j \) are independent, for \( i \neq j \). In other word, \( \epsilon \sim N(0, \sigma) \); \( E(\epsilon) = 0 \) and \( cov(\epsilon, \epsilon) = 0 \) , for \( i \neq j \).

OLS method is minimize sum square and will result:

\[ \hat{\beta} = (X'X)^{-1}(X'Y) \]  \hspace{1cm} (2)

and

\[ V(\hat{\beta}) = (X'X)^{-1}\sigma^2 \]  \hspace{1cm} (3)

with \( X'X \) is nonsingular matrix. If \( X'X \) singular, then estimator of \( \beta \) can be find by general inverse matrix. The estimator is not unique and the general solution (Kshirsagar, 1983) is:

\[ \hat{\beta} = \hat{\beta} + (I - H)\zeta \]  \hspace{1cm} (4)

With \( H \) is idempotent matrix size \( pxp \), \( \zeta \) is any vector , meanwhile:

\[ \hat{\beta} = S^{-1}X'Y \]  \hspace{1cm} (5)

where \( S \) is general invers of \( S = X'X \).

3. System Equations
Assume there are \( M \)-equations that are related because the error terms are correlated. This system of \( M \) ‘seemingly unrelated regression equations’ can be written in matrix format as follows.

\[ y_1 = X_1\beta_1 + \mu_1 \]
\[ y_2 = X_2\beta_2 + \mu_2 \]
\[ \vdots \]
\[ y_M = X_M\beta_M + \mu_M \]

Using more concise notation, this system of \( M \)-equations can be written as

\[ y_i = X_i\beta_i + \mu_i \]  \hspace{1cm} for \( i = 1, 2, \ldots, M \)  \hspace{1cm} (6)
Where y is a Tx1 column vector of observations on the ith dependent variable; X is a TxK matrix of observations for the K-I explanatory variables and a column vector of 1’s for the ith equation (i.e., the data matrix for the ith equation; β is the Kx1 column vector of parameters for the ith equation; and µ is the Tx1 column vector of disturbances for the ith equation.

To combine the M-equations into one single large equation, stack the vectors and matrices and can be written more concisely as:

\[
y = X\beta + \mu
\]  

Where y is a (M•T)x1 column vector of observations on the dependent variables for the M-equations; X is a (M•T)x(M•K) matrix of observations on the explanatory variables; with the columns of 1’s, for the M-equations; β is a (M•K)x1 column vector of parameters for the M-equations; and µ is a (MxM) column vector of disturbances for the M-equations. The M•Tx1 disturbance vector in (6) and (7) is assumed to have the following variance-covariance matrix

\[
V(\mu) = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM}
\end{bmatrix} \otimes I
\]

\[
V(\mu) = \Sigma \otimes I
\]  

where Σ is an MxM matrix of variances and covariances for the M individual equations.

4. Aitken’s Generalized Least Squares (GLS) Estimator
In a formal sense we now regard (6) and (7) as a single-equation regression model. To apply Aitken’s GLS we pre-multiply both sides of (7) by a matrix H which is such that

\[
E(H\mu\mu'H') = H\Sigma H' = I
\]  

In terms of transformed variables, the original variables pre-multiplied by H, the system is now satisfies the usual assumptions of the OLS. Thus application of OLS will yield a best linear unbiased estimator which is:

\[
\beta^* = (X'HHX)^{-1} X'HHy = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1} y
\]
Let $H' H = \Omega^{-1}$ which implies $H \Omega^{-1} H' = I$. In fact, several such matrices $H$ exist, so that we can assume $H = H'$. Now, according to [3], to derive the form of best linear unbiased estimator of $\beta$ for generalized regression model under assumption that $\Omega$ is known, define:

\[
y^* \equiv Hy
\]
\[
X^* \equiv HX
\]

By the usual mean-variance calculations,
\[
E(y^*) = HX\beta = X^* \beta
\]
And
\[
V(y^*) = HV(y)H' = H(\sigma^2 \Omega)H' = \sigma^2 I
\]

Since $X^* = HX$ is clearly nonrandom with full column rank if $X$ satisfies assumption, the classical regression model applies to $y^*$ and $X^*$, so the Gauss-Markov theorem implies that the best linear unbiased estimator of $\beta$ is

\[
\hat{\beta}_{GLS} = (X^* X^*)^{-1} X^* y
\]
\[
= (X' \Omega^{-1} X')^{-1} X' \Omega^{-1} y
\]

But since this estimator is also linear in the original dependent variable $y$, it follows that this GLS estimator is best linear unbiased using $y$ [3]. Also, the usual estimator of the scalar variance parameter $\sigma^2$ will also be unbiased if $y^*$ and $X^*$ are used:

\[
s^2_{GLS} = \frac{1}{N-K} \left[ y^* - X^* \hat{\beta}_{GLS} \right]' \left( y^* - X^* \hat{\beta}_{GLS} \right)
\]
\[
= \frac{1}{N-K} \left[ y - X \hat{\beta}_{GLS} \right]' \Omega^{-1} \left( y - X \hat{\beta}_{GLS} \right)
\]

has $E(s^2_{GLS}) = \sigma^2$ by the usual arguments.

If $y$ is assumed multinormal, $y \approx N(X\beta, \sigma^2 \Omega)$, then the existing results for classical OLS imply that $\hat{\beta}_{GLS}$ is also multinormal $\hat{\beta}_{GLS} \approx \left( \beta, \sigma^2 (X' \Omega^{-1} X)^{-1} \right)$ and is independent of $s^2_{GLS}$, with

\[
\frac{(N-K)s^2_{GLS}}{\sigma^2} \approx \chi^2_{N-K}.
\]

If the sample data are generated by the seemingly unrelated regression model, then the GLS estimator is unbiased, efficient, and the maximum likelihood estimator. The reason the GLS estimator is more precise than the OLS estimator is that it uses the information about the nonspherical disturbances contained in $W$ to obtain estimates of the parameters.

The GLS estimator is not a feasible estimator, because you don’t know the elements of the variance-covariance matrix of disturbances, $W$, for the big equation. More detailed discussion of the GLS theory can also be found in e.g. [4] and [5].

5. The Application

Aitken’s GLS method will be applied and compared to OLS method on demand function of animal protein in Indonesia. Data used in this study is household consumption/expenditure data collected by Central Bureau of Statistics in 2016 on 36 provinces in Indonesia [6], [7]. There are three equations will be estimated that represent three type of animal protein, those are fish, meat, and egg/milk.
To get estimated disturbance covariance matrix conveniently, it can write:

\[
\hat{\Sigma} = \begin{bmatrix} \hat{\beta}_1 & 0 & 0 \\ 0 & \hat{\beta}_2 & 0 \\ 0 & 0 & \hat{\beta}_3 \end{bmatrix}
\]

Or \( Y = X\hat{\beta} + \mu \)

Then \( \mu'\mu = (Y - X\hat{\beta})(Y - X\hat{\beta}) = Y'Y - \hat{\beta}'X'X\hat{\beta} \)

\[ (13) \]
\[
\begin{bmatrix}
  y_1'y_1 & y_1'y_2 & y_1'y_3 \\
  y_2'y_1 & y_2'y_2 & y_2'y_3 \\
  y_3'y_1 & y_3'y_2 & y_3'y_3 \\
\end{bmatrix}
\begin{bmatrix}
  \hat{\beta}_1'X_1'X_1 & \hat{\beta}_1'X_1'X_2 & \hat{\beta}_1'X_1'X_3 \\
  \hat{\beta}_2'X_2'X_1 & \hat{\beta}_2'X_2'X_2 & \hat{\beta}_2'X_2'X_3 \\
  \hat{\beta}_3'X_3'X_1 & \hat{\beta}_3'X_3'X_2 & \hat{\beta}_3'X_3'X_3 \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
  0.001401 & -0.000843 & -0.000367 \\
  -0.000843 & 0.001156 & -0.000409 \\
  -0.000367 & -0.000409 & 0.000749 \\
\end{bmatrix}
\]

Which is equal to \((T-5) \{ \hat{\mu}_1 \} \) where \(T=36\), the number of observations on each variable. To obtain Aitken’s coefficient estimates, multiply the last matrix into the following vector:

\[
\begin{bmatrix}
  23.126 \\
  7.471 \\
  \ldots \\
  -44.764 \\
  \ldots \\
  -14.508 \\
  -0.899 \\
  \ldots \\
  32.519 \\
  \ldots \\
  -6.727 \\
  -6.479 \\
  \ldots \\
  8.37 \\
\end{bmatrix}
\]

The point estimates so obtained along with their estimated variances are shown in Table 1.
Tabel 1. Parameter Estimate of Demand Function Using GLS and OLS Method

| Dependent Variable | Independent Variable | GLS Method | OLS Method |
|--------------------|----------------------|------------|-----------|
|                    |                      | Coefficient estimate | Variance of coefficient estimator | Coefficient estimate | Variance of coefficient estimator |
| Y, (fish)          | X,                   | 0.746       | 0.010816  | 0.746       | 0.010816  |
|                    | X,                   | 0.241       | 0.003969  | 0.241       | 0.003969  |
|                    | X,                   | 0.588       | 0.004225  | 0.588       | 0.004225  |
|                    | X,                   | -0.049      | 0.002209  | -0.049      | 0.002209  |
|                    | X,                   | -1.444      | 0.034225  | -1.444      | 0.034225  |
| Y, (meat)          | X,                   | -0.468      | 0.009025  | -0.468      | 0.009025  |
|                    | X,                   | -0.029      | 0.003364  | -0.029      | 0.003364  |
|                    | X,                   | -0.428      | 0.003481  | -0.428      | 0.003481  |
|                    | X,                   | -0.057      | 0.001764  | -0.057      | 0.001764  |
|                    | X,                   | 1.049       | 0.028224  | 1.049       | 0.028224  |
| Y, (egg/milk)      | X,                   | -0.217      | 0.005776  | -0.217      | 0.005776  |
|                    | X,                   | -0.209      | 0.002116  | -0.209      | 0.002116  |
|                    | X,                   | -0.09       | 0.002304  | -0.09       | 0.002304  |
|                    | X,                   | 0.169       | 0.001156  | 0.169       | 0.001156  |
|                    | X,                   | 0.270       | 0.018225  | 0.270       | 0.018225  |

It can be seen that the parameter estimate of GLS and OLS methods are equal because the assumption of normality and heteroscedasticity aren’t violated. The model is very good as shown as 99% coefficient of determination (output of SAS program). The test the normality of distribution of error at 5% level show that the error term has normal multivariate distribution. Meanwhile, the scatterplot of the residual and prediction show the random pattern as below.

Figure 1. Scatterplot of Prediction and Residual of demand function of fish, meat, egg/milk

The White test of constant variance at 5% significance level show that the residual have constant variance.

6. Conclusion
Aiken’s generalized least square (GLS) can be applied in the analysis data provided by a single cross-section household consumption/expenditure study when several commodities are to be estimated with correlated disturbance error. The method yields coefficient estimators which efficient than single
equation least square estimators with 99% coefficient of determination. The test the normality show that the error term has normal multivariate distribution. The White show residual have constant variance.

7. Concluding Remarks
It can be noted that the point estimates yielded by the two methods differ. This is to be expected since different quadratic forms are minimized in the two approaches and also if one method is more efficient in another, the estimates yielded by the two methods cannot always or even usually be identical as [8]. What it is important to realize is that it makes good sense to use the Aitken quadratic form. In this form we have weighted deviations; that is, the data in the sample are not all given the same weight but are weighted by elements of the covariance matrix’s inverse.

8. Acknowledgement
We thank Nia Kusniati, SSi. for assistance with data processing and Prof. Dr. I Wayan Mangku, MSc from Departement of Mathematics, Bogor Agriculture University for comments that improved the manuscript.

References
[1] Virgantari F, Daryanto A, Harianti H, Kuntjoro S U 2011 Analisis Permintaan Ikan di Indonesia: Pendekatan Model Quadratic Almost Ideal Demand System (Quaids) Jurnal Sosial Ekonomi Kelautan dan Perikanan 6 (2) pp 191-203
[2] Zellner A 1962 An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests forAggregation Bias Journal of the American Statistical Association 57 298 pp 348-368
[3] Seddighi H R, Lawler K A and Katos A V 2000 Econometrics (London: Routledge)
[4] Amemiya T 1977 The Maximum Likelihood and the Nonlinear ThreeStage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model Econometrica Journal of the Econometrics Society 45 (4) pp 955-968
[5] Greene W H 2003 Econometric Analysis 5th edition (New York: Prentice-Hall)
[6] Hakiki G 2016 Consumption of Calorie and Protein of Indonesia Population and Province Book 2 (Jakarta: Badan Pusat Statistik)
[7] Supriyanto S 2016 Consumption Expenditure of Population of Indonesia by Province Book 3 (Jakarta: Badan Pusat Statistik)
[8] Virgantari F, Wijayanti H, Multina L B 2014 Kajian Model Almost Ideal Demand System pada Fungsi Permintaan Pangan Hewani di Indonesia Prosiding Seminar Nasional Matematika 2014 Universitas Indonesia Jakarta