DOI <https://doi.org/10.18551/rjoas.2017-03.11>

RANK-BASED ESTIMATION FOR ASYMMETRIC PRICE TRANSMISSION MODELLING

Acquah De-Graft Henry, Associate Professor

Department of Agricultural Economics and Extension, University of Cape Coast,
Cape Coast, Ghana

E-mail: henrydegratacquah@yahoo.com

ABSTRACT

This paper introduces and compares Rank-based estimation and the Ordinary Least Squares (OLS) methods for modelling the Granger and Lee asymmetric price transmission model when the true data generating process is known. Monte Carlo simulation results indicate that the estimates of the coefficients of the asymmetric price transmission model derived from the Least squares and the Rank-based estimation methods are accurate and equivalent or close to their true values for normal data regardless of variability in sample size. For data with outliers, Least squares method is affected by outliers and yields inaccurate estimates of the coefficients of the asymmetric price transmission model across various sample sizes. Rank-based estimation remains robust to outliers in large samples and provides estimates of the coefficients of the asymmetric price transmission model that are accurate and nearly equivalent to their true values. The evidence from Monte Carlo experimentation suggests that the proposed Rank-based estimation is likely to do no worse than the OLS with normal dataset and promise to do better when the dataset has outliers within the asymmetric price transmission modelling context.

KEY WORDS

Monte Carlo Simulation, Rank-based Estimation, Granger and Lee Asymmetry, Ordinary Least Squares Estimation, Outlier.

An econometric methodology to investigate asymmetric adjustments that relies on ordinary least squares (OLS) estimation is proposed by Granger and Lee (1989). Consequently, the OLS estimation of the Granger and Lee Asymmetric model has been widely employed in modelling price asymmetry. However, the OLS estimation is only optimal under certain assumptions. For example, if the normality assumption of the error term and other assumptions of the linear regression model are not violated. But under circumstances commonly encountered with economic data such as contaminated data (data with outliers), some underlying assumptions for which the OLS is efficient is violated.

In the presence of outliers, the OLS may produce wrong asymmetric adjustment coefficient, which forms the basis for detecting asymmetry. In effect, wrong conclusions will be made from the Granger and Lee asymmetric model and its variants when underlying assumptions of the OLS are violated due to the presence of outliers. For example, Douglas (2010) finds the presence of outliers in price data in asymmetric price transmission analysis. He detects asymmetry in the contaminated data but on dropping the outliers, finds no evidence of asymmetry. Kind (2015) also notes that data used in agricultural price analysis from developing countries are more often found to have outliers and as result, estimates obtained from such contaminated data lose their value. In a spatial price analysis of selected agricultural commodities, Karfakis and Rapsomanikis (2007) noted that the error correction modelling methods are not robust to the presence of outliers in the price series.

An alternative approach to estimate the Granger and Lee Asymmetric model whilst concurrently accommodating outliers in the data is to employ Rank-based regression. Rank-based regression remains robust to the presence of outliers. This approach has been met with success in estimation of linear models in the presence of outliers as detailed in Jureckova (1971); Jaeckel (1972); McKean and Hettmansperger (1978); and Kloke and McKean (2015).

Rank-based regression estimation provides valuable information even if the data is contaminated with outliers as commonly encountered in price data used in asymmetric price transmission analysis. However, less effort has been made to compare the Rank-based estimation and OLS methods for asymmetric price transmission analysis in the presence of outliers. The purpose of this research is therefore to investigate by use of Monte Carlo methods, the performance of the Rank-based regression and OLS methods in estimating the Granger and Lee asymmetry using data with and without outliers.

The paper is planned as follows. The introduction is followed by the methodology which discusses the Granger and Lee Asymmetric model, Ordinary Least Squares (OLS) and Rank-based Regression. The results and discussion present Monte Carlo simulations of Granger and Lee asymmetric model and demonstrates the ability of Ordinary Least Squares and Rank-based regression to estimate true values of Granger and Lee asymmetric data generating process. Finally, the paper ends with a conclusion.

METHODS OF RESEARCH

Granger and Lee Asymmetry. Granger and Lee (1989) Asymmetric Error Correction Model data generating process can be specified as follows:

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2^+ (y - x)_{t-1}^+ + \beta_2^- (y - x)_{t-1}^- + \varepsilon_{1,t} \quad \varepsilon_{1,t} \sim N(0, \delta^2) \quad (1)$$

If y and x are variables integrated of the order one processes [I (1)] that are cointegrated, then there exists a long run equilibrium relationship between y and x which is captured by a symmetric error correction term $(y-x)$. Asymmetric adjustments can be introduced by decomposition of the error correction term into positive and negative components as follows:

$$(y - x)_t^+ = \begin{cases} (y - x)_t, & \text{if } (y - x)_t > 0 \\ \text{zero} & \text{otherwise} \end{cases} \quad (2)$$

$$(y - x)_t^- = \begin{cases} (y - x)_t, & \text{if } (y - x)_t < 0 \\ \text{zero} & \text{otherwise} \end{cases} \quad (3)$$

Asymmetry is incorporated by allowing the speed of adjustment to differ for the positive and negative components of the Error Correction Term (ECT) since the long run relationship captured by the ECT is symmetric. The test for symmetry in eq. (1) is conducted by determining whether the coefficients $(\beta_2^+ \text{ and } \beta_2^-)$ are identical (that is $H_0 : \beta_2^+ = \beta_2^-$). The Granger and Lee Asymmetric Error correction model in eq. (1) can be considered as a standard regression model as specified below:

$$y = X\beta + \varepsilon_t \quad \varepsilon_t \sim iidN(0, \sigma^2) \quad (4),$$

Where: y and X are a response and explanatory variables with the explanatory variables defined to include asymmetric adjustment terms. Consequently, the estimation of parameters of the Granger and Lee Asymmetric model can be done using Ordinary Least Squares method and Rank-based technique. Empirical results are computed for the two techniques and compared.

Ordinary Least Squares Estimation (OLS). Suppose a data which is made up of n observations $\{Y_i X_i\}_{i=1}^n$. Each observation has an outcome variable Y_i and a vector of p explanatory variables X_i . Here, the outcome variable is a linear function of the explanatory variables in a linear regression model. In this regard, the general model can be of the form:

$$Y = f(X_1, X_2, X_3) + \varepsilon \quad (5)$$

This could be re-written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \quad (6),$$

Where: β_i , $i=1, 2, 3$ are unknown parameters.

Eq. (6) is written in a matrix form as:

$$Y_i = X_i \beta + \epsilon_i \quad (7)$$

Where $y = (y_1, \dots, y_n)^T$, $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$, $\beta = (\beta_1, \dots, \beta_n)^T$ and $X = \begin{pmatrix} 1 & X_{11} & x_n \\ \dots & \dots & \dots \\ 1 & x_{n1} & x_{nn} \end{pmatrix}$

The column of the ones incorporates the intercept term. Estimation of β from eq. (7) which involves minimizing the sum of square errors as follows:

$$\sum \epsilon_i^2 = \epsilon^T \epsilon = (y - X\beta)^T (y - X\beta) \quad (8)$$

Differentiating with respect to β , setting it to zero, we find that $\hat{\beta}$ satisfies:

$$X^T X \hat{\beta} = X^T y \quad (9)$$

The estimated β is the Best Linear Unbiased Estimators (BLUE) under Gauss-Markov theorem, and can be estimated using OLS. The OLS method of estimating β can be estimated by $\hat{\beta}$ which is given by the explicit formula:

$$\hat{\beta} = (X'X)^{-1} X'y \quad (10)$$

The matrix $(X'X)^{-1} X'$ is referred to as the Moore-Penrose pseudo inverse matrix of X . After estimation of β , the fitted values (or predicted values) from the regression will be:

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1} X'y \quad (11)$$

Here $\hat{\beta}$ is unbiased and has variance $(X'X)^{-1} \sigma^2$.

From the above, estimating the variance (σ^2), we find that $E\hat{\epsilon}^T \hat{\epsilon} = \sigma^2(n-p)$. This suggests that the estimator $\hat{\sigma}^2 = \frac{\hat{\epsilon}^T \hat{\epsilon}}{n-p} = \frac{RSS}{n-p}$ is an unbiased estimator of σ^2 . Where $n-p$ is the degree of freedom of the model.

It is assumed that the errors are independent and identically distributed (iid) with mean 0 and variance σ^2 , thus $\varepsilon \sim N(0, \sigma^2 I)$. Now since $y = X\beta + \varepsilon$, implies that $y \sim N(X\beta, \sigma^2 I)$, which is a compact description of the regression model.

Rank-Based Regression. Consider a linear model of the form:

$$Y_i = \alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + e_i \text{ for } i = 1, \dots, n. \quad (12)$$

From eq. (12), the reduced form can be stated as:

$$Y_i = \alpha + x_i^T \beta + \varepsilon_i \text{ for } i = 1, \dots, n. \quad (13),$$

Where: Y_i is a continuous response variable, x_i is vector of explanatory variable, α is intercept parameter, β is vector of coefficients, and ε_i is the error term. The ε is assumed to be iid with a continuous pdf function, $f(x)$.

Rearranging eq. (13) in matrix notation can be stated as:

$$Y = \alpha 1 + X\beta + e \quad (14),$$

Where: $Y = [Y_1, \dots, Y_n]^T$ is $n \times 1$ vector of response variable, $X = [X_1, \dots, X_n]^T$ is $n \times p$ design matrix, and $e = [e_1, \dots, e_n]^T$ is $n \times 1$ vector of error terms.

Comparing the least square estimator and the rank estimator, the least squares estimator minimizes the Euclidean distance between Y and $\hat{Y}_{LS} = X\hat{\beta}_{LS}$. The R estimator on the other hand is obtained through the Jaeckel's (1972) dispersion function, given by:

$$D(\beta) = \|Y - X\beta\|_{\varphi}, \quad (15)$$

$D(\beta)$ is a convex function of β and provides a robust measure of distance between Y and $X\beta$. $\|\cdot\|_{\varphi}$ is a pseudo-norm defined as:

$$\|\mu\|_{\varphi} = \sum_{i=1}^n a(R(u_i))u_i \quad (16)$$

In eq. (16), scores are generated as $a(i) = \varphi\left(\frac{i}{n+1}\right)$, and φ is a nondecreasing score function which is defined on interval $(0, 1)$.

The R estimator of β is defined as:

$$\hat{\beta}_{\varphi} = \text{Argmin } \|Y - X\beta\|_{\varphi}. \quad (17)$$

Eq. (17) is consistent with asymptotically normal distribution given by:

$$\hat{\beta}_{\varphi} \sim N(\beta, \tau_{\varphi}^2 (X^T X)^{-1}), \quad (18),$$

Where: τ_{φ} is the scale parameter. Eq. (18) follows a normal distribution with the variance-covariance matrix:

$$V_{\hat{\beta}_{\varphi}} = \begin{bmatrix} K_n & -\tau_{\varphi}^2 \bar{x}' (X'X)^{-1} \\ -\tau_{\varphi}^2 (X'X)^{-1} \bar{x} & \tau_{\varphi}^2 (X'X)^{-1} \end{bmatrix},$$

Where: $K_n = n^{-1}\tau_S^2 + -\tau_{\varphi}^2 \bar{x}' (X'X)^{-1} \bar{x}$. The vector $\bar{x}$ is the vector of column averages of X and τ_S is the scale parameter.

RESULTS AND DISCUSSION

Comparison of Rank-based and Ordinary Least Squares Estimation. The Granger and Lee asymmetric error correction model data generating process is specified as follows:

$$\Delta y_t = 0.7 + 0.5x_t - 0.25(y_t - x_t)^+_{t-1} - 0.75(y_t - x_t)^-_{t-1} + \varepsilon \quad (19),$$

y_t and x_t are generated as non-stationary variables that are integrated of the order one. There exists a cointegrating relationship between y and x which is defined by the error correction term $(y_t - x_t)_{t-1}$. The positive and negative components of the error correction term are denoted by $(y_t - x_t)^+_{t-1}$ and $(y_t - x_t)^-_{t-1}$. The errors are generated from a normal distribution with a mean 0 and a variance of 1 [$\varepsilon \sim (0,1)$] for normal data. For the data with outliers, nine observations of the errors generated for the normal data with values generated from the normal distribution with a mean of 0 and a variance of 1 is replaced with nine observations from the normal distribution with a mean of 20 and variance of 1 ($\varepsilon \sim N(20,1)$).

In order to investigate the performance of the Rank-based estimators and OLS in estimating the true values of the asymmetric price transmission model, 1000 regressions based on the Granger and Lee model specified in eq. (19) is estimated. The Monte Carlo experimentation is conducted under conditions of different sample sizes (50, 150 and 500) and asymmetry given by $(\beta_2^+, \beta_2^-) \in (\beta_2, \beta_3) \in (-0.25, -0.75)$ for the normal data as well

as the data with outliers. This study assigns the asymmetric adjustment parameters (β_2^+, β_2^-) in the spirit of Cook et al (1999, 2000) and Acquah (2012, 2013).

The results obtained from the Monte Carlo analysis for the normal data are reported in Table 1. Results of 1000 Monte Carlo simulations indicate that the estimates of the coefficients of the asymmetric price transmission model derived from the Rank-based analysis are accurate and close to their true parameter values for the data without outliers (Normal data) with small and moderate sample sizes (50 and 150). The estimates of the coefficients of the asymmetric price transmission model derived from the least squares methods are accurate and equal to their true parameter values for the data without outliers (Normal data) with small and moderate sample sizes (50 and 150). Noticeably, the estimates of the coefficients of the asymmetric price transmission model derived from the Least squares method and the Rank-based analysis are accurate and equivalent to their true parameter values for normal data with large sample size (500).

In summary, Table 1 illustrates that in the absence of outliers, the OLS and Rank-based analysis performed well, with the averaged estimates all nearly equivalent or close to their true values of $\beta_0 = 0.7, \beta_1 = 0.5, \beta_2 = -0.25, \beta_3 = -0.75$ regardless of differences in sample sizes.

The results are consistent with Chen, Tang, Lu and Tu (2014) who noted that in the absence of outliers, OLS and Rank-based methods performed well, with the averaged estimates all nearly identical to the true values in linear regression analysis. Similarly, Ryan (1997) notes that robust methods such as Rank-based estimation methods perform almost as well as the OLS when the data is free from mistakes and influential data points.

Table 1 – Normal Data (Without Outliers)

Sample Size	Properties of Data	Method	Estimates			
			β_0	β_1	β_2	β_3
N=50	Normal	OLS	0.70	0.50	-0.25	-0.75
		Rank-Estimation	0.70	0.50	-0.25	-0.76
Sample Size	Properties of Data	Method	Estimates			
			β_0	β_1	β_2	β_3
N=150	Normal	OLS	0.70	0.50	-0.25	-0.75
		Rank-Estimation	0.70	0.50	-0.25	-0.74
Sample Size	Properties of Data	Method	Estimates			
			β_0	β_1	β_2	β_3
N=500	Normal	OLS	0.70	0.50	-0.25	-0.75
		Rank-Estimation	0.70	0.50	-0.25	-0.75

Based on 1000 Monte Carlo Simulation.

The results obtained from the Monte Carlo analysis for the data with outliers are reported in Table 2. Results of 1000 Monte Carlo simulations indicate that the estimates of the coefficients of the asymmetric price transmission model derived from the Rank-based analysis are accurate and close to their true parameter values for the data with outliers in large sample (500). Noticeably, as sample size increase from small through moderate to large sample, estimated coefficients of the asymmetric price transmission model move closer to their true parameter values in the Rank-based regression analysis.

In the presence of outliers, the least squares method performed poorly as shown in Table 2. In small, moderate and large samples of 50, 150 and 500 respectively, the ordinary least squares (OLS) estimator performs poorly with its parameter estimates entirely different from the true parameter values of $\beta_0 = 0.7, \beta_1 = 0.5, \beta_2 = -0.25, \beta_3 = -0.75$ as defined in the data generating process.

From the foregoing discussion, it is obvious that the results of the Rank-based analysis is similar to the least squares and close to their true values in the data without outliers. However, in data with outliers, the least squares is affected by outliers in small, moderate and large samples whilst the Rank based analysis remains robust to outliers in large samples.

Table 2 – Data with Outliers

Sample Size	Properties of Data	Method	Estimates			
			β_0	β_1	β_2	β_3
N=50	With Outliers	OLS	3.05	0.50	-0.16	-1.08
		Rank-Estimation	0.85	0.50	-0.24	-0.78
Sample Size	Properties of Data	Method	Estimates			
			β_0	β_1	β_2	β_3
N=150	With Outliers	OLS	1.00	0.47	-0.44	-2.01
		Rank-Estimation	0.71	0.50	-0.26	-0.87
Sample Size	Properties of Data	Method	Estimates			
			β_0	β_1	β_2	β_3
N=500	With Outliers	OLS	1.00	0.50	-0.26	-0.52
		Rank-Estimation	0.72	0.50	-0.25	-0.73

Based on 1000 Monte Carlo Simulation.

The results are consistent with Ryan (1997) assertion that robust methods such as Rank-based estimation methods perform much better than OLS when the data has outliers. Similarly, Chen, Tang, Lu and Tu (2014) noted that in the presence of outliers, classic linear models yield extremely large estimates that are un-interpretable, whilst in contrast, the rank regression model generated estimates close to their true values. Atikatu (2015) also in an empirical study showed that rank-based methods were more robust in estimation when the distribution of the error term of the dataset was non-normal and also in the presence of outlying observations, whilst the OLS method was very non-robust. Results from the study indicated that 5% outlier-contamination was enough to cause some instability in the estimates of the OLS method.

CONCLUSION

The performance of Rank-based estimators have been investigated in asymmetric price transmission regression modelling. The findings suggest that the Rank-based estimation yield similar results as the OLS with normal data. However, when outliers are present in the data, the least squares does not provide correct estimates of the coefficients of the asymmetric price transmission model in small, moderate and large samples of data. Rank-based estimation on the other hand is robust and provides correct estimates of the coefficients of the asymmetric price transmission model in large samples. The results of the simulation indicate that the Rank-based estimation can be considered an alternative to the OLS technique in asymmetric price transmission estimation and may yield accurate results in large samples when the data contains outliers.

REFERENCES

1. Acquah, H. D. (2012). A bootstrap approach to testing for symmetry in the Granger and Lee Asymmetric Error Correction Model. RJOAS, 11(11), 33-36.
2. Acquah, H. D. (2013). Using bootstrap method to evaluate the power of tests for non-linearity in asymmetric price relationship. Journal of Economics and Behavioral Studies, 5, (4), 237-241.

3. Atikatu, A. (2015). Rank-based adaptive method of estimating beta. Thesis submitted to the department of mathematics, Kwame Nkrumah University of Science and Technology
4. Chen, T., Tang, W., Lu, Y., & Tu, X. (2014). Rank regression: an alternative regression approach for data with outliers. *Shanghai archives of psychiatry*, 26(5), 310.
5. Cook, S., Holly, S., & Turner, P. (1999). The Power of tests for non-linearity: the case of Granger–Lee asymmetry, *Economics Letters*, 62, pp.155–159.
6. Cook, S., Holly, S., & Turner, P. (2000). The Power of Tests for Non-linearity: The Escribano-Pfann Model, *Computational Economics*, 15, pp. 223-226.
7. Douglas, C. C. (2010). Do gasoline prices exhibit asymmetry? Not usually!. *Energy Economics*, 32(4), 918-925.
8. Granger, C. W. J., & Lee, T. H. (1989). Investigation of production, sales and inventory relationships using multicointegration and non-symmetric error correction models. *Journal of applied econometrics*, 4(S1), S145-S159.
9. Jaeckel, L. A. (1972). Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals. *The Annals of Mathematical Statistics*, 43,1449-1458.
10. Jureckova, J. (1971) Non-parametric Estimate of Regression Coefficients. *The Annals of Mathematical Statistics*, 42,1328-1338.
11. Karfakis, P., & Rapsomanikis, G. (2007). Margins across time and space: threshold cointegration and spatial pricing applications to commodity markets in Tanzania. In workshop on staple food trade and market policy options for promoting development in eastern and southern Africa.
12. Kind, M (2015). Analysis of Market Integration- An Alternative Approach. MSc Thesis submitted to the Agricultural Economics and Rural Policy Group, Wageningen University, Netherland.
13. Kloeke, J. & McKean, J. W. (2015). *Nonparametric Statistical Methods Using R*. New York: CRC Press. ISBN-13:978-1-4398-7343-4.
14. McKean, J. W., & Hettmansperger, T. A (1978). Robust Analysis of the General Linear Model Based on one Step r-estimates. *Biometrika*, 65(3):571.
15. Ryan, T. P. (1997). *Modern regression methods*. New York, NY: John Wiley & Sons, Inc.