DOI <https://doi.org/10.18551/rjoas.2017-05.05>

THE EFFECT OF OUTLIERS ON THE PERFORMANCE OF AKAIKE INFORMATION CRITERION (AIC) AND BAYESIAN INFORMATION CRITERION (BIC) IN SELECTION OF AN ASYMMETRIC PRICE RELATIONSHIP

Acquah De-Graft H., Associate Professor

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

E-mail: henrydegratacquah@yahoo.com

ABSTRACT

Asymmetric price transmission modelling aims to select one model that best captures the asymmetric data generating process from a set of competing models using model selection methods. However, such an interest in model selection outpace an awareness that outliers in data can have a disproportionate impact on model ranking. In order to explore the issue, the effect of outliers on the performance of commonly used Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) in selection of asymmetric price relationship are evaluated under conditions of different sample size. Monte Carlo experimentation indicated that the ability of the model selection methods to identify the true asymmetric price relationship decreases with an increase in outliers in moderate and large samples. With 5% outlier-contamination in large samples, both AIC and BIC fail to identify the true asymmetric price relationship. BIC outperforms AIC in selecting the asymmetric data generating process in large samples with outliers. However, in small samples, the effect of outliers on the performance of AIC and BIC in selection of the correct asymmetric model remains unclear.

KEY WORDS

Model selection, Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC), asymmetry, Monte Carlo, outliers.

Asymmetric price transmission modelling aims to select one model that best captures the asymmetric data generating process (DGP) from a set of competing models. This entails choosing the model that provides the best fit to the data on the basis of information criteria. The justification for this approach is that the model providing the best fit is the one that closely approximates the underlying asymmetric data generating process. However, such an interest in model selection outpace an awareness that outliers in data can have a disproportionate impact on model ranking. In the presence of outliers, a model may provide the best fit among a set of competing models without necessarily closely approximating the data generating process.

Outliers are observations that are rare and for some reasons, different from majority of the observations (see Barnett & Lewis, 1994 for a detail discussion). The effect of outliers on the ranking of competing models has been mentioned several times in the statistics literature (e.g., Hoeting, Raftery, & Madigan, 1996; McCann, 2006; Ronchetti, 1997; Ronchetti, Field, & Blanchard, 1997). Previous research on Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) has found them sensitive to outliers in regression analysis (e.g., Atkinson & Riani, 2008; Chik, 2002; Laud & Ibrahim, 1995; Le, Raftery, & Martin, 1996). Atkinson and Riani (2008) notes that the sensitivity of model selection indices, such as AIC, to outliers is an often overlooked issue.

Little is known about the relative performance of different information criteria in asymmetric price transmission modelling when the data contains outliers. Empirically, less effort has been made in examining the influence of outliers on model selection within the asymmetric price transmission modelling context. Notably, the ability of the commonly used model selection methods (AIC and BIC) to select the true asymmetric price transmission model in the presence of outliers has not yet been extensively investigated and is not well understood. An important question which remains unanswered is how well will AIC and BIC

perform when outliers are present in the data used for price transmission analysis. In the presence of outliers, will AIC and BIC point to the correct asymmetric price transmission model?

In order to address this issue, this paper empirically evaluate and compare the performance of the two commonly used model selection criteria, AIC and BIC in choosing between alternative methods of testing for asymmetry in the presence of outliers. The paper contributes towards understanding the effect of outliers on the model selection performance of AIC and BIC in asymmetric price transmission modelling framework. The true data generating process is known in all experiments and the Monte Carlo simulations are essential in deriving the model recovery rates of the true model.

METHODOLOGY OF RESEARCH

The process of selecting a statistical model from a set of candidate models is called model selection. Information criteria provides the basis of choosing between competing models. The basic concept of information criteria is to select statistical models that simplify description of the data and model. In effect, information criteria emphasizes minimising the amount of information required to express the data and model.

In addressing the problem of choosing among competing models, information criteria allows one to select the model that gives the most accurate description of the data. It addresses the trade-off between descriptive accuracy and minimizing the number of parameters.

Akaike Information Criteria (AIC). A widely used information criteria, Akaike Information Criteria (AIC) was introduced by Akaike (1973; 1974) via Kullback Lieber divergence. AIC is an estimate of the relative expected Kullback Lieber distance of a given model from the true model. AIC is derived as an asymptotically unbiased estimator of the expected Kullback-Liebler discrepancy between the true and a fitted model. It is defined as:

$$AIC = -2 \log(L) + 2p \quad (1),$$

Where: the first term $-\log(L)$ is the negative maximum log-likelihood of the data given the model parameter estimates and the second term p , is the number of parameters in the model. AIC aims to find the best approximating model to the data generating process. Models with smallest AIC values are deemed as best.

Bayesian Information Criteria (BIC). Another widely used information criteria, the Bayesian Information Criteria (BIC) was proposed by Schwarz (1978) as an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model. The computation of the BIC is based on the empirical log-likelihood of the candidate model and does not require the specification of priors. BIC is defined as:

$$BIC = -2 \log(L) + p \log(n) \quad (2),$$

Where: n is the sample size, p is number of parameters in the model and $-\log(L)$ is the negative maximum log-likelihood of the data given the model. BIC is consistent and tends to select the true model with a probability of one as sample size increases. Under this selection criteria, models with minimum BIC are preferred.

Asymmetric Price Transmission Models. Several econometric models have been developed to estimate asymmetric price transmission. They include the Houck (1977) model (HKM), Standard Error Correction Model (SECM) and the Complex Error Correction Model (CECM). For the purpose of brevity, the standard asymmetric error correction model, the complex asymmetric error correction model and the Houck's model are denoted by SECM, CECM and HKM respectively. The Houck's model is specified as follows:

$$\Delta y_t = \beta_1^+ \Delta x_t^+ + \beta_1^- \Delta x_t^- + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (3)$$

The Houck's Model (HKM) relates changes in the response price (Δy_t) to the positive and negative changes in the other price ($\Delta x_t^+, \Delta x_t^-$). Dynamic variants of this model can be estimated to distinguish between short and long run asymmetries.

The Standard Error Correction Model (SECM) is specified as follows:

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2^+ (y - x)_{t-1}^+ + \beta_2^- (y - x)_{t-1}^- + \varepsilon_t \quad \varepsilon_t \sim N(0, \delta^2) \quad (4)$$

The Standard Error Correction Model relates changes in response price (Δy_t) to changes in the other price (Δx_t) as well as changes in the Error Correction Term (ECT) which is decomposed into positive and negative components ($(y - x)_{t-1}^+, (y - x)_{t-1}^-$). Eqn. (4) was proposed in Granger and Lee (1989).

Engle and Granger (1987) notes that if y and x are cointegrated, then an error correction representation exist. Cointegration is first established by estimating the long run relationship between price y_t and x_t . The lagged residuals from the Eqn. (5) denotes the Error Correction Term and is included in the standard error correction model.

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (5)$$

The contemporaneous response term (Δx_t) is segmented in Von Cramon-Taubadel and Loy (1996). This leads to the following specification in which contemporaneous and short-run response to departures from the cointegrating relation are asymmetric if $\beta_1^+ \neq \beta_1^-$ and $\beta_2^+ \neq \beta_2^-$ respectively:

$$\Delta y_t = \beta_1^+ \Delta x_t^+ + \beta_1^- \Delta x_t^- + \beta_2^+ (y - x)_{t-1}^+ + \beta_2^- (y - x)_{t-1}^- + \varepsilon_t \quad (6)$$

$$\varepsilon_t \sim N(0, \delta^2)$$

In this case, a joint F-test can be used to determine symmetry or asymmetry of the price transmission process. Notably, asymmetries specified affects the direct impact of price increases and decreases as well as adjustments to the equilibrium level. Where Δx_t^+ and Δx_t^- are the positive and negative changes in x_t and the remaining variables are defined as in the standard error correction model. The asymmetric ECM with complex dynamics nests the Houck's model in first difference.

RESULTS AND DISCUSSION

In order to illustrate the ability of the model selection methods to recover the true model when the data contains outliers, a series of Monte Carlo simulation experiments are conducted and the results are reported below. The simulation is based on the Standard Error Correction Model (SECM) data generating process specified as follows:

$$\Delta y_t = 0.7 \Delta x_t - 0.25 (y_t - x_t)_{t-1}^+ - 0.75 (y_t - x_t)_{t-1}^- + \varepsilon \quad (7),$$

y_t and x_t are generated as I (1) non-stationary variables that are cointegrated. The error correction terms ($(y_t - x_t)_{t-1}^+, (y_t - x_t)_{t-1}^-$) represent the positive and negative deviations from the long run equilibrium relationship between y_t and x_t . For data without outliers, the

errors are generated from a normal distribution with a mean 0 and a variance of 1 ($\varepsilon \sim N(0,1)$). In order to create outliers in the data, various percentages of outliers (0, 2, 3, 4 and 5 percent) are introduced into the data without outliers. For example, two percent of the number of observations of the errors generated for the normal data with values generated from a normal distribution with a mean of 0 and a variance of 1, were replaced with two percent of the number of observations from the normal distribution with a mean of 20 and variance of 1 ($\varepsilon \sim N(20,1)$) for a chosen sample size. This is repeated for 3, 4 and 5 percent of outliers given the various sample sizes 50, 150 and 500 respectively. The data generating process is simulated 1000 times with different percentages of outliers and across different sample sizes. For each simulation, the ability of the model selection methods to recover the true data generating process is evaluated. The percentage of samples in which each competing model provides a better model fit than the other competing models is referred to as the model recovery rates. The recovery rates are derived using 1000 Monte Carlo simulations. In effect, the amount of samples in which each model fits better than the other competing models is measured out of the 1000 samples and expressed as a percentage. Thus, the values obtained from each model selection criteria are calculated as the arithmetic mean based on 1000 samples.

Generally, the overall power of the model selection methods (AIC and BIC) to select the true data generating process decreased with increase in the percentage of outliers. Noticeably, outliers have a substantial impact on selection power in moderate and large samples. In effect, the ability of AIC and BIC to select the true asymmetric data generating process decreased with increase in the percentage of outliers in moderate and large samples.

Table 1 – Relative performance of the model selection methods (Small Sample)

% of Outliers	Experimental Criterion		Model fitted			
			Methods	CECM (%)	HKM (%)	SECM (DGP) (%)
0	n=50	$\sigma = 1$	AIC	16.2	5.1	78.7
			BIC	5.2	13	81.8
2	n=50	$\sigma = 1$	AIC	16.2	5.1	78.7
			BIC	5.2	13	81.8
3	n=50	$\sigma = 1$	AIC	18.5	3.3	78.2
			BIC	6.8	9.5	83.7
4	n=50	$\sigma = 1$	AIC	17.2	4.1	78.7
			BIC	6.5	10.9	82.6
5	N=50	$\sigma = 1$	AIC	18.5	3.3	78.2
			BIC	6.8	9.5	83.7

Recovery rates based on 1000 replications.

Comparison of the different criteria are illustrated in Tables 1, 2, and 3. Model recovery rates are presented for each criteria under various sample size conditions with varying percentage of outliers. In small samples, the performance trends of the model selection criteria remains unclear. Notably in small samples, outliers have an unclear effect on the ability of the model selection methods to recover the correct model. This is because in small samples, the expected number of outliers is n times the percentage of outliers. This leads to a few outliers since n is small. The outlier effect becomes unclear in small samples and more pronounced in large samples. For outliers' percentages of 0, 2, 3, 4 and 5, the ability of the model selection methods to choose the true asymmetric data generating process is seriously distorted in small samples. For example, without outliers in the data ($n=50$), AIC and BIC recovered 78.7% and 81.8% respectively of the true asymmetric data generating process. However, with introduction of 5% outliers into the data ($n=50$), AIC and BIC recovered 78.2% and 83.7% respectively. In small sample size of 50, the model recovery rates derived when the data contains outliers does not show any substantial difference from those recovery rates derived when there was no outlier in the data.

Table 2 – Relative performance of the model selection methods (Moderate Sample)

% of Outliers	Experimental Criterion		Model fitted			
			Methods	CECM (%)	HKM (%)	SECM (DGP) (%)
0	n=150	$\sigma = 1$	AIC	15.2	0.1	84.7
			BIC	2.5	0.1	97.4
2	n=150	$\sigma = 1$	AIC	15.6	0.0	84.4
			BIC	3.2	0.0	96.8
3	n=150	$\sigma = 1$	AIC	14.4	6.3	79.3
			BIC	3.1	20.4	76.5
4	n=150	$\sigma = 1$	AIC	14.5	18.8	66.7
			BIC	2.3	47	50.7
5	n=150	$\sigma = 1$	AIC	14.2	38.2	47.6
			BIC	0.7	71.6	27.7

Recovery rates based on 1000 replications.

In moderate sample size of 150, a clear pattern is seen in the performance of the model selection methods to recover the true asymmetric data generating process as the percentage of outliers' increases. For example, as the percentage of outliers increase from 0 to 5%, there is a substantial decline in the ability of the model selection methods to select the true asymmetric data generating process, as the recovery rates for AIC and BIC decrease from 84.7% and 97.4% to 47.6% and 27.7% respectively. Noticeably, the performance of AIC and BIC in the selection of asymmetric price relationship is affected by outliers. Similarly, previous studies (Laud & Ibrahim, 1995) found that AIC and BIC are not robust to outliers or influential data points. Ronchetti, Field and Blanchard (1997) also observed that outliers have an undue influence on model chosen.

Table 3 – Relative performance of the model selection methods (Large Sample)

% of Outliers	Experimental Criterion		Model fitted			
			Methods	CECM (%)	HKM (%)	SECM (DGP) (%)
0	n=500	$\sigma = 1$	AIC	15.4	0.0	84.6
			BIC	1.6	0.0	98.4
2	n=500	$\sigma = 1$	AIC	25.7	0.1	74.2
			BIC	3.8	2.4	93.8
3	n=500	$\sigma = 1$	AIC	40.1	3.2	56.7
			BIC	5.6	24.3	70.1
4	n=500	$\sigma = 1$	AIC	46.8	9.8	43.4
			BIC	4.0	46.1	49.9
5	n=500	$\sigma = 1$	AIC	49.5	16.1	34.4
			BIC	4.1	57.2	38.7

Recovery rates based on 1000 replications.

In large samples of 500, outliers have a more pronounced effect on asymmetric price transmission model selection performance. This is because the effect of outliers become more pronounced in large samples. Subsequently, as sample size increases, the number of outliers in the sample increases. This is because in large samples, the expected number of outliers is n times the percentage of outliers. This increases the number of outliers since n is large. For example, the recovery rates of AIC and BIC decreases from 84.6% and 98.4% to 34.4% and 38.7% respectively when outlier percentages in the data increases from 0 to 5%. With 5% outlier contamination, both AIC and BIC performs poorly and fail to identify the true asymmetric model in large samples. Similarly, McCann (2006) noted that standard model selection methods such as AIC and BIC performs poorly when the data contains outliers.

Under the influence of 5% outlier contamination in large samples, AIC selected the complex asymmetric model against the true data generating process. Similarly, BIC selected the simpler asymmetric model against the true data generating process. BIC outperforms AIC in selecting the suitable asymmetric price relationship in large samples with outliers.

Similarly, in a model selection analysis using posterior probabilities, Le, Raftery and Martin (1996) note that BIC outperforms AIC for data with outliers.

CONCLUSION

The study examined the ability of AIC and BIC to clearly identify the true asymmetric data generating process in the presence of outliers in the data. Generally, the Monte Carlo simulation results indicate that the ability of AIC and BIC to clearly identify the correct model among competing models decreases with increases in outliers in the data.

Under unstable conditions such as small sample size, the effect of outliers on the performance of AIC and BIC in selection of the correct asymmetric model remains unclear. However, in moderate and large samples, there is a persistent decline in the performance of AIC and BIC to recover the true asymmetric data generating process. As the percentage of outliers increase to 5% outlier contamination, both AIC and BIC fail to identify the true asymmetric model in large samples.

The comparison provided contributes to knowledge and understanding of the effects of outliers on the relative performance of AIC and BIC in an asymmetric price transmission modelling framework. The study contributes to the literature on asymmetric price transmission modelling by making researchers aware of the failure of AIC and BIC to select the correct asymmetric price transmission model when outlier percentages in data (large samples) are high. Investigation into the performance of robust modifications of AIC and BIC in selection of asymmetric price transmission models in the presence of outliers represents a fruitful avenue for future research.

REFERENCES

1. Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. 2nd International Symposium on Information Theory, Budapest. pp. 267-281.
2. Akaike, H. (1974). A new look at the Statistical Model Identification. IEEE Transactions on Automatic Control, 19:716-723
3. Atkinson, A. C., & Riani, M. (2008). A robust and diagnostic information criterion for selecting regression models. Journal of the Japanese Statistical Society, 38:3-14.
4. Barnett, V. & Lewis T. (1994). Outliers in statistical data, 3rd ed., John Willey, Chichester.
5. Chik, Z. (2002). The effect of outliers on the performance of order selection criteria for short time series. Pakistan Journal of Applied Sciences, 2:912-915.
6. Engle, R.F., & Granger, C.W.J. (1987). Co-integration and error correction: Representation, estimation and testing. Econometrica, 55:251-276.
7. 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 4:135-159.
8. Hoeting, J., Raftery, A. E., & Madigan, D. (1996). A method for simultaneous variable selection and outlier identification in linear regression. Computational Statistics and Data Analysis, 22:251-270.
9. Houck, J. P. (1977). An Approach to specifying and estimating nonreversible Functions, American Journal of Agricultural Economics, 59:570-572.
10. Laud, P. W., & Ibrahim, J. G. (1995). Predictive model selection. Journal of the Royal Statistical Society, Series B, 57:247-262.
11. Le, N. D., Raftery, A. E., & Martin, R. D. (1996). Robust Bayesian model selection for autoregressive processes with additive outliers. Journal of the American Statistical Association, 91:123-131.
12. McCann, L. (2006). Robust model selection and outlier detection in linear regression. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.
13. Ronchetti, E. (1997). Robustness aspects of model choice. Statistica Sinica, 7:327-338.
14. Ronchetti, E., Field, C., & Blanchard, W. (1997). Robust linear model selection by cross-validation. Journal of the American Statistical Association, 92:1017-1023.
15. Schwarz, G. (1978). Estimating the Dimension of a Model. Annals of Statistics, 6:461-464.

16. Von Cramon-Taubadel, S., & Loy, J.-P. (1996). Price Asymmetry in the international Wheat Market: Comment. *Canadian Journal of Agricultural Economics*, 44, pp. 311-317.