Electronic Journal of Applied Statistical Analysis
Vol. 17, Issue 01, March 2024, 153-171
DOI: 10.1285/i20705948v17n1p153

Hybrid Methodology for Sparse Selection of Generalized Estimating Equations Model for the Drivers of Firm Value

Robert Nyamao Nyabwanga 1*[a]

[a]*Kisii University, Mathematics and Actuarial Science Department, 408-40200, Kisii, Kenya*

15 March 2024

The study proposes a two-step hybrid methodology for sparse generalized estimation equations modeling of the drivers of shareholder value creation. Through the methodology, the validity of the Gordon constant growth model is established and other non-dividend factors' contribution to shareholder value creation is assessed. The two-step hybrid method involves picking out the right intra-subject correlation matrix and set of regressors using EAIC and QIC respectively (EAIC-QIC) and then obtaining the penalized GEE estimators of the selected model. Penalization is useful in removing redundant regressors from the final model. The performance of the proposed method was compared to that of exclusively using QIC method in selecting both the correlation matrix and set of regressors. The study results showed that, whereas EAIC preferred the parsimonious order one auto-aggressive {AR(1)} structure for the data, QIC preferred the unstructured matrix which estimates the highest number of correlation parameters. Using the AR(1) structure and Algorithm 2, the GEE model chosen had higher efficiency compared to when QIC is used to select both the correlation matrix and regressors. Based on the results, the study concludes that adopting hybrid methods enhances efficiency of GEE estimators. On firm value, the study concludes that besides the elements in the Gordon-Constant growth model, the financial health of a firm is a vital indicator of value creation ability by firms.

*Corresponding authors: nyamaonyabwanga@gmail.com

©Università del Salento
ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

Keywords: Generalized Estimating Equations, Model Selection, Correlation Structure, Sparsity, Penalized GEE, Shareholder Value Creation.

1 Introduction

The Constant-Growth Model (Gordon, , 1959) which was built on the assumption that the company is a going concern and will grow forever, paying dividends at a continually increasing rate, has been the fundamental model in predicting value creation by firms for their shareholders in decades. Based on the model, value creation is a function of dividends and the discount rate:

$$P_0 = \frac{D_1}{r-g} = \frac{D_0(1+g)}{r-g} \tag{1}$$

where, P_0 =the present-day market value of the anticipated flows of dividends per share(D_1); D_0=Current Earning Per Share; g = growth in earnings and r is the equity cost of capital. According to model 1, dividend per share and economic profitability proxied by income per share are the key determinants of the present market value of a share. This leads to the model:

$$M_v = \alpha_0 + \alpha_1 D + \alpha_2 Y \qquad (Gordon,, \ 1959). \tag{2}$$

Where M_v= the year-end price, D = the year's dividend per share, and Y = the year's income per share.

Given that growth in earnings (g) is a function of the return on equity (ROE) and retention ratio (γ) and further that earnings per share is a function of ROE and the book value of equity shares (B_v), model (1) can be modified to take the form;

$$\frac{M_v}{B_v} = \frac{ROE - g}{r - g} \tag{3}$$

Considering the ratio the current year's dividend and the book value of equity (d) and equation (3) it can be observed that ROE adjusted for the cost of equity capital (ROE-r) which is a measure of economic profitability (E_p), growth rate of earnings and dividend policy (d) are the key drivers of firm value hence Equation (2) can be refined to take the form:

$$\frac{M_v}{B_v} = \alpha_0 + \alpha_1 E_p + \alpha_2 g + \alpha_3 E_P * g + \alpha_4 d + \epsilon \tag{4}$$

Model (4) implies that the key drivers of shareholder value creation are growth in earnings (g), economic profitability proxied by earnings per share (γ) and dividend policy.

The consideration of dividend policy proxied by the dividend pay-out ratio as a driver for firm value was also considered by Hansda, et al. (2020) and Agung et al. (2021a). Hansda, et al. (2020) using dynamic panel regression with two-step system Generalised Method of Moments (GMM) established that dividend policy had no significant effect on firm value while Agung et al. (2021a) just like Bataha et al. (2023) who both used multiple linear regression method, established that dividend policy had a significant positive effect on firm value which were in line with the signalling theory. These contrasting findings limits the development of theory on the relationship between dividend policy and firm value. The disagreements have been attributed to changes in study contexts, research periods, research design and methods of analysis. For instance the methods employed by Hansda, et al. (2020), Bataha et al. (2023) and Agung et al. (2021a)) resulted to differing results.

Further to the factors considered in Equation (4), the study considers the Modigliani and Miller (MM) hypothesis which under some specified assumptions established that debt policy had no effect on firm value and include debt policy as a factor that can influence value creation by firms. Chen and Chen, (2011) considered debt policy represented by leverage (proxied by the debt-equity ratio) and established a negative relationship with firm value, a result that was in contrast to the MM hypothesis.

Likewise, the study considered accounting profitability measured by return on assets (ROA) ratio, firm size, financial distress measured by the Atlman Z score and working capital policy (WCP). Pandey, (2005) established that accounting profitability was an important value driver. In relation to firm size, Chen and Chen, (2011) found out that firm value had a positive relationship with profitability while Mule et al. (2015) showed that firm size had no significant relationship with firm value. Honjo and Harada, (2006) considered the number of board members (Bsize), working capital policy and financial health as indicators of firm value. Just like Honjo and Harada, (2006), Nguyen and Faff, (2007) established that there was no significant relationship between firm value and board size. The study therefore considers the expanded model in Equation (5)

$$\begin{aligned} \frac{M_v}{B_v} &= \alpha_0 + \alpha_1 g + \alpha_2 E_p + \alpha_3 E_p * g + \alpha_4 d + \alpha_5 ROA + \alpha_6 Lev + \alpha_7 Fsize \\ &\quad + \alpha_8 Bsize + \alpha_9 Z + \alpha_{10} WCP + \epsilon \end{aligned} \tag{5}$$

The standard multiple regression model has often been used in establishing the relationship between the various factors and firm value (Agung et al., 2021a; Bataha et al., 2023). This method does not take into consideration the intra-class correlation and assumes that observations within and between clusters are independent. As observed by Gordon, (1959), a cross-section of studies that used regressions analysis yielded high correlation but with differing regression coefficients and corresponding standard errors amongst samples from different industries. This has led to questions being raised on the economic significance of the results despite the fact that the variation in price among shares is of paramount importance in guiding investors' choice of investment possibilities and in directing corporate financial policy formulation.

This study sought to extend the use of the method of generalized estimating equations (GEE) to corporate finance so as to consider each firm as a cluster from which a number of measurements are taken and have a defined working correlation structure that defines the within-cluster correlation. The method will also allow for the choice of both an appropriate link function that depend on the distribution of the response variable and an appropriate set of covariates for the mean structure (Carey and Wang , 2011). GEE is a population average method developed based on the quasi-likelihood theory hence does not need one to specify the distribution of the response variable but only the mean and variance functions of the response observations (Carey and Wang , 2011). Further, as observed by Cui and Qian (2007), even with a mis-specified working correlation structure, GEE analysis still yields consistent regression coefficient estimators.

Pan (2001) developed and championed for the routine use of Quasi-likelihood Information Criteria (QIC) for the selection of both the correlation matrix and best subset of explanatory variables. He commended the QIC criteria for its good performance in variable selection. QIC has however been established to more often select a wrong correlation structure leading to less efficient GEE estimators to the extent of 40%. The finding on correlation matrix selection performance by Pan (2001) were supported by other findings which established that QIC was weak in picking out the true correlation matrix for repeated measurements (Nyabwanga et al., 2019a; Wang et al. , 2012; Cui and Qian , 2007). The wanting performance of QIC has over the years led to several modifications by scholars such as Hin and Wang, (2009) who developed the Correlation Information Critieria (CIC), Chen and Lazar (2012) who developed EAIC and EBIC among others in efforts to increase chances of selecting the correct matrix hence enhance efficiency of the estimates. On their part Oyebayo and Mohdi, (2019) championed for the use of Hybrid methodology since there was no single method that could effectively select both the correlation matrix and set of covariates. In line with their recommendation, the study proposes a Two-Step Hybrid methodology and applies it in modelling the drivers of firm value for firms listed in the Nairobi Securities Exchange (NSE), Kenya. Further, the performance of the proposed method is compared to the QIC-only benchmark method.

The rest of the paper is organized as follows: Section 2 on Materials and Methods provides data description, a review of the GEE method in the context of the drivers of shareholder value creation, GEE model selection in which the Hybrid methodology is presented in Algorithms 1 and 2 and finally the efficiency measures used. Section 3 presents the results and discussion while section 4 provides the conclusions.

2 Materials and Methods

2.1 Data Description

Data for the study was collected for a sample of 53 firms out of 61 listed in the NSE that were in operation as at January 2012. The representative sample was obtained using proportionate stratified Sampling method. The corresponding number of firms per cluster were: 8 agricultural; 3 auto-mobile and accessories; 10 banking; 8 commercial services; 4 construction and allied ; 3 energy and petroleum; 5 insurance; 3 investment; 8 manufacturing and 1 telecommunications. Data collected covered a period of 6 years (2012-2017)which resulted to 318 binary outcomes for $\frac{Mv}{Bv}$ and all the covariates captured in equation (5).

2.2 The GEE Model

For each firm i, let $Y_{it} \in (0,1)$ be the value creation history, $i = 1 \cdots, 53$ and $t = 1, \cdots, 6$ such that:

$$Y_{it} = \begin{cases} 1 & \text{if } \frac{MV}{BV} > 1 \\ 0 & \text{if } \frac{MV}{BV} \leq 1 \end{cases}$$

Therefore, $Y_i = \{y_{i1}, \cdots, y_{im}\}^T$ is the $m \times 1$ random vector of value creation history for the i^{th} firm, $i = 1, \cdots, n$. Further, let $\boldsymbol{X_{jit}} = (X_{1it}, X_{2it}, \cdots, X_{kit})$ be the vector of explanatory variables where$k = 1, 2, \cdots, 10$, $t = 1, 2, \cdots, m$, X_1=economic profitability, X_2=growth in earnings, X_3=interaction between growth in earnings and economic profitability, X_4=logarithm of total assets, X_5=leverage (debt-equity ratio), Return on Assets(ROA), X_6=Dividend payout Ratio, X_7=Level of fincial health calculated using the Altman's Z score given by $Z = 6.56\frac{W_c}{T_a} + 3.26\frac{R_E}{T_a} + 6.72\frac{EBIT}{T_a} + 1.02\frac{M_v}{T_L}$ where WC=Working Capital; RE=Retained Earnings; EBIT=earnings before interest and tax; T_L=Total value of Liabilities and T_a=Total assets, X_8= Number of Board members, X_9= ROA, X_{10}=Liquidity ratio representing the working capital policy.

For the set of data $\{Y_{it}, \boldsymbol{X_{jit}}\}$, let $E(Y_{it}|\boldsymbol{X_{jit}}) = \mu_{it}$ relate to $\boldsymbol{X_{jit}}$ through an appropriate link function so that

$$g(\mu_{it}) = \boldsymbol{X_{jit}^T}\boldsymbol{\beta} \tag{6}$$

$$Var(Y_{it}|\boldsymbol{X_{jit}}) = \phi V(\mu_{it}) \tag{7}$$

where $\mu_{it} = Pr(Y_{it} = 1|\boldsymbol{X_{jit}})$ represents the chances of a firm i creating value for its stockholders at time t hence $(1 - \mu_{it})$ is the probability that it fails to create value for its shareholders at time t and

$$Var(Y_i) = \begin{pmatrix} C_{(1,1)} & C_{(1,2)} & \cdots & C_{(1,m)} \\ C_{(2,1)} & C_{(2,2)} & \cdots & C_{(2,m)} \\ \vdots & \vdots & \ddots & \vdots \\ C_{(m,1)} & C_{(m,2)} & \cdots & C_{(m,m)} \end{pmatrix} \tag{8}$$

Where $C_{(t,t)}$,$t = 1, \cdots, 6$ are variances and the off-diagonal elements are covariances. $\boldsymbol{\beta} = \{\beta_0, \beta_1, \cdots, \beta_{10}\}^T$ is a $p \times 1$ vector of regression coefficients such that;

$$\mu_{it} = g^{-1}(\beta_0 + \sum_{j=1}^{k} \boldsymbol{\beta_j X_{jit}}) \tag{9}$$

and considering that Y_{it} is a binary response, the logit link function can be applied hence equation (9) can be re-written as:

$$log(\frac{\mu_{it}}{1 - \mu_{it}}) = \beta_0 + \sum_{j=1}^{k} \boldsymbol{\beta_j X_{jit}} \tag{10}$$

$$\mu_{it} = \frac{Exp(\beta_0 + \sum_{j=1}^{k} \boldsymbol{\beta_j X_{jit}})}{1 + Exp(\beta_0 + \sum_{j=1}^{k} \boldsymbol{\beta_j X_{jit}})} \tag{11}$$

Liang and Zeger (1986), established that by solving the system of generalized estimating equations in Equation (12) iteratively, the vector of parameters $\boldsymbol{\beta}$ are obtained.

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^T} V_i^{-1}(y_i - \mu_i) = 0 \tag{12}$$

where

$$V_i = A^{\frac{1}{2}} R_i(\rho) A^{\frac{1}{2}} \tag{13}$$

is the model based variance-covariance matrix for Y_i and $R(\rho)$ is the working correlation matrix for cluster i whose order is $m_i \times m_i$ fully specified by ρ.

$$A_i = \begin{pmatrix} \sigma^2_{(Y_{i1})} & 0 & \dots & 0 \\ 0 & \sigma^2_{(Y_{i2})} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2_{(Y_{im})} \end{pmatrix} \tag{14}$$

Remark 2.1. Since GEE is not a likelihood based method of estimation, a solution to equation (12) may be established by using the Iterative Weighted Least Squares in which the estimating equations are solved by linearizing μ_i around an initial estimate say $\boldsymbol{\beta^0}$ and also evaluating V_i at the same $\boldsymbol{\beta^0}$. Let $h(\mu_i) = \eta_i = \boldsymbol{X}^T_{\boldsymbol{jit}} \boldsymbol{\beta_j}$ such that;

$$\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} = \frac{\partial \mu_i}{\partial \eta_i} \times \frac{\partial \eta_i}{\partial \boldsymbol{\beta}} \times \frac{\partial \mu_i}{\partial \eta_i} \boldsymbol{X}_i$$

then, by first-order Taylor approximation of μ_i in the neighborhood of $\mu_i^{(0)}$

$$\begin{aligned} \mu_i &= \mu_i^{(0)} + \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} (\boldsymbol{\beta} - \boldsymbol{\beta}^0) \\ &= \mu_i^{(0)} + \frac{\partial \mu_i}{\partial \eta_i} \boldsymbol{X}_i^T \{\boldsymbol{\beta} - \boldsymbol{\beta}^0\} \end{aligned} \tag{15a}$$

and

$$y_i - \mu_i = y_i - \mu_i^0 - \frac{\partial \mu_i}{\partial \eta_i} \boldsymbol{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}^0) \tag{15b}$$

Plugging (15a) and (15b) into (12) results into:

$$\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \eta_{\ i}} V_i^{-1} \boldsymbol{X_i} (y_i - \mu_i^0 - \frac{\partial \mu_i}{\partial \eta_i} \boldsymbol{X}_i^T (\boldsymbol{\beta} - \boldsymbol{\beta}^0)) = 0 \tag{16}$$

This solves for $\boldsymbol{\beta}$ in the next iterate hence the updating formula is given as;

$$\boldsymbol{\beta}^{(1)} = (\boldsymbol{X}^T W \boldsymbol{X}^{-1}) \boldsymbol{X}^T W_\kappa \tag{17}$$

Where $\boldsymbol{X}$ is a matrix of covariates, $W = \{\frac{\partial \eta_i}{\partial \mu}_i\}^2 V_i$ and κ is the adjusted dependent variable such that $\kappa^1_{(1)} = h(\mu_i^{(0)} + \frac{\partial \eta_i}{\partial \mu}_i (y_i - \mu_i^{(0)}))$. This procedure will produce a sequence of estimates $\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(21)} \cdots, \boldsymbol{\beta}^{(t)}$ and the iterations are stopped when $||\hat{\boldsymbol{\beta}}^{(t+1)} - \boldsymbol{\beta}^{(t)}|| = \varepsilon$, where ε is the set threshold value.

Lemma 2.2. *For a given $\boldsymbol{\beta}$ from equation (12), let $\lambda_i = \sum_{t=1}^{m} y_{it}$ be the total number of times a firm creates value in the 'm' financial years, then $\lambda_i \sim Binomial(m, P_i(\hat{\boldsymbol{\beta}}))$ and the probability that a firm i creates value for the shareholders at least once given $\boldsymbol{X_j}$ is;*

$$Pr(Y_{it} = 1|\lambda_i \geq 1) = 1 - \{1 - P_i(\hat{\boldsymbol{\beta}})\}^m \tag{18}$$

Such that,

$$\mu_{it} = E(Y_{it} = 1|\lambda_i \geq 1) = \frac{P_{it}(\hat{\boldsymbol{\beta}})}{1 - \prod_{t=1}^{m}(1 - P_{it}(\hat{\boldsymbol{\beta}}))} \tag{19a}$$

$$Var(Y_{it}|\lambda_i \geq 1) = \frac{P_{it}(\hat{\boldsymbol{\beta}})\{1 - P_{it}(\hat{\boldsymbol{\beta}}) - \prod_{t=1}^{m}(1 - P_{it}(\hat{\boldsymbol{\beta}}))\}}{[1 - \prod_{t=1}^{m}(1 - P_{it}(\hat{\boldsymbol{\beta}}))]^2} \tag{19b}$$

Theorem 2.3. *If $log(\frac{\mu_{it}}{1-\mu_{it}}) = \beta_0$ such that $\mu_{it} = g^{-1}(\beta_0) = P_0$, the estimating equation for β_0 is:*

$$\sum_{i=1}^{n}\{\lambda_i - \frac{mP_0}{1 - (1 - P_0)^m}\} \tag{20}$$

Proof. Assuming the independence correlation matrix, then, $D_i = \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^T} = A_i\boldsymbol{X_i}$. From equation (12) we have;

$$\begin{aligned} U(\boldsymbol{\beta}) &= \sum_{i=1}^{n}(A_i\boldsymbol{X_i})^T V_i^{-1}(y_i - \mu_i) = 0 \\ &= \sum_{i=1}^{n}(\boldsymbol{X_i})^T A_i A_i^{-1}(y_i - \mu_i) = \sum_{i=1}^{n}(\boldsymbol{X_i})^T(y_i - \mu_i) = 0 \\ &= \sum_{i=1}^{n}\{1, \cdots, 1\} \times \left(\begin{pmatrix} E(Y_{i1}) \\ E(Y_{i2}) \\ \vdots \\ E(Y_{im}) \end{pmatrix} - \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{im} \end{pmatrix}\right) \\ &= \sum_{i=1}^{n}\lambda_i - \sum_{t=1}^{m}\mu_{it} = 0 \end{aligned} \tag{21}$$

However, based on the results in Equation (19a) and since $P_{it} = g(\beta_0) = P_0$, then $\mu_{it} = \frac{P_0}{1-(1-P_0)^m}$. Hence the estimating equations for β_0 can be expressed as:

$$\sum_{i=1}^{n}\{\lambda_i - \frac{mP_0}{1 - (1 - P_0)^m}\}$$

which ends the proof. □

2.3 Shareholder Value creation GEE Model Selection

In GEE, model selection centers on selecting a working correlation structure $R(\rho)$ and a suitable set of covariates. The study proposes the use of empirical likelihood based AIC (EAIC) proposed by Chen and Lazar (2012) in choosing the correct correlation matrix and QIC in selecting the covariates.

2.3.1 Two-Step EAIC-QIC Hybrid Methodology

The proposed hybrid methodology is implemented using Algorithms 1 and 2 presented below in which its proposed that the best correlation matrix be first selected and then followed by the choice of the best subsets of covariates with some regularization that will ensure that only the informative set of regressors is selected.

STEP I:Selection of Intra-Class Correlation Matrix

Algorithm 1 Application of EAIC for Intra-Class Correlation Matrix Selection

if $\Re^f$ represent the Empirical Likelihood Ratio(ELR) defined based on the full model **then**;

Formulate the full model ELR in terms of $\boldsymbol{\rho}$ and $\boldsymbol{\beta}$ where $\boldsymbol{\rho} = \{\rho_1, \cdots, \rho_{m-1}\}$ whose maximization is taken with respect to the probabilities $p_1, \cdots, p_n$.

$$\begin{aligned} \Re^f(\boldsymbol{\beta}) &= Sup\{\prod_{i=1}^{n} np_i | p_i \geq 0, \sum_{i-1}^{n} p_i = 1, \sum_{i-1}^{n} p_i g(\boldsymbol{X_i^T}\boldsymbol{\beta}) = 0\} \\ g(\boldsymbol{X_i^T}\boldsymbol{\beta}) &= \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^T} V_i^{-1}(\boldsymbol{\mu})(y_i - \mu_i) \\ V_i^{-1}(\boldsymbol{\mu}) &= A_i^{-0.5} R^{-1}(\rho) A_i^{-0.5} \end{aligned} \tag{22}$$

if there are 's' correlation structures considered, say $R_1(\boldsymbol{\rho}), \cdots, R_S(\boldsymbol{\rho})$ **then**

[i] define the ELR of the GEE estimator based on each $R_s, s = 1 \cdots S$ i.e. $\Re^f_{R_1}, \cdots, \Re^f_{R_S} = \Re^f\{\beta^{R_1}, \hat{\rho}^{R_1}\}, \cdots, \Re^f\{\beta^{R_S}, \hat{\rho}^{R_S}\}$.

[ii] For each R_s, obtain the $MELE(\hat{\boldsymbol{\beta}}^s)$. This is the same as the GEE estimator $(\boldsymbol{\beta}^s_G)$ in Equation (22).

for each of the s models **do**;

[i] Compute the EAIC values of based on the formula by Chen and Lazar (2012) i.e.

$$EAIC_s = -2log\Re^f(\hat{\theta}^s_{GEE}) + 2dim(\theta^s)$$

where the s candidate models based on the s intra-class correlations are parameterized by $\theta^s(s = 1, \cdots, S)$ and $(\hat{\theta}^s_{GEE}) = \binom{\hat{\boldsymbol{\beta}}^s_{GEE}}{\hat{\rho}^s_{GEE}}$ is the GEE estimator based on the correlation matrix $R_s(\rho)$. $\hat{\rho}^s_{GEE}$ is the method of moment estimator of ρ given $\hat{\boldsymbol{\beta}}^s_{GEE}$ and R_s

[ii] Select the best correlation matrix for the data. i.e. $R^s_{best}(\rho) = argmin(EAIC_s)$

end for

end if

end if

Remark 2.4. Use of the full model to select the best correlation matrix ($R^s_{best}(\rho)$) is plausible as it contains most information to predict the outcome variable.

STEP II: Selection of Covariates

Algorithm 2 Application of QIC for Covariates Selection

for 'k' regressor variables and $R^s_{best}(\rho)$ from Algorithm 1 **do**

[i] Generate using the "Dredge" procedure in Multi-Model Inference (MuMIn) R package, all the $2^k, p = k+1$ GEE models (this excludes the constant only model).

[ii] $\forall$, $M_i, i \in \mathbb{Z}, i = 1, 2, \cdots 2^k$ models, compute the QIC values based on the formula;

$$QIC^{R^s_{best}}(\rho) = -2Q\{\hat{\beta}^{R^s_{best}}(\rho)|(Y_i, X_{ij})\} + 2tr(\hat{\Omega}_I \hat{V}_{R^s_{best}}(\rho)) \tag{23}$$

Where, $\hat{\Omega}_I = \frac{1}{n}\sum_{i=1}^n (\frac{\partial \mu_i}{\partial \beta^T})^T V_i^{-1} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^T}$ is the model based variance-covariance matrix and $\hat{V}_{R_{s_{best}}(\rho)} = \hat{\Omega}_I^{-1}\{(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}^T})^T V_i^{-1}(y_i - \mu_i)(y_i - \mu_i)^T V_i^{-1} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^T}\}\hat{\Omega}_I^{-1}$ is the robust variance estimator under the best working correlation matrix (R^s_{best})

for Each of $M_i, i = 1, \cdots, 2^k$ **do**

[i] Rank the models based on the criteria suggested by Burnham and Anderson, (2002)

[ii] extract models whose $\Delta_i \leq 2$ where $\Delta_i = QIC_{M_i} - QIC_{min}$. In this case the best model is $\{M_i : \Delta_i = 0\}$.

for The the best model, $\{M_i : \Delta_i = 0\}$ **do**

Re-fit using Penalized GEE to obtain pnalized estimators ($\boldsymbol{\beta_j}$) and the corresponding standard errors. PGEE employs the SCAD penalization to further remove uninformative regressors.

end for

end for

end for

Remark 2.5. Fitting the final using PGEE is informed by findings by Nyabwanga et al. (2019b) that QIC has high over-fitting probabilities hence part (4) in Algorithm 1 is meant to ensure that non-informative regressors are removed from he final selected model so as to reduce model variability while keeping the model bias at minimum.

Definition 2.6. Let $U(\hat{\boldsymbol{\beta}})$ denote the original GEE estimates and $S(\hat{\boldsymbol{\beta}})$ the PGEE estimates. Then;

$$S(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^T} V_i^{-1}(y_i - \mu_i) - q'_\lambda(|\boldsymbol{\beta}|) \circ sign(\boldsymbol{\beta}) \tag{24}$$

where $q_\lambda(|\boldsymbol{\beta}|) = \{q_\lambda(|\beta_1|), \cdots, q_\lambda(|\beta_p|)\}^T$ is a p-dimensional vector of penalty functions and $sign(\boldsymbol{\beta}) = \{sign(\beta_1), \cdots, sign(\beta_p)\}^T$ with $sign(t) = I(t > 0) - I(t < 0).q'_\lambda(|\boldsymbol{\beta}|) \circ sign(\boldsymbol{\beta})$ denotes the Hadamard product of the two vectors.

The smoothly clipped absolute deviation (SCAD) penalty by Fan and Li, (2001), is employed owing to its ability to simultaneously achieve unbiasedness, sparsity and continuity. The SCAD penalty function is zero for lager coefficients and is relatively large if β_j is close to or equal to zero hence will guard against under-fitting. For more details on SCAD penalization, see Fan and Li, (2001). To obtain the SCAD regularized estimates, the study employed the Minorization-maximization algorithm together with the Newton-Raphson algorithm. See Wang et al. (2012).

2.4 Relative Efficiency of the Proposed Hybrid method

To establish the relative efficiency (RE) of the new procedure compared to the standard QIC only procedure, the study employed V-fold Cross-Validation to obtain the MSE of each procedure in which

$$MSE_v = \frac{1}{v}\sum_{i\in v}(y_i - \hat{y}_i)^2$$

and the V-fold cross-validation estimate is the average of the V estimates of the test errors $MSE_i, \cdots, MSE_v$ where

$$CV_v = \frac{1}{V}\sum_{i}^{V} MSE_i$$

and

$$RE = \frac{MSE(\hat{\beta}_{GEE}^{QIC})}{MSE(\hat{\beta}_{GEE}^{EAIC-QIC})}$$

If $RE > 1$, GEE estimates under the Hybrid method have higher efficiency than its counterpart procedure. The GEE estimates' efficiency for the new hybrid procedure will be lesser compared to the exclusive use of QIC If $RE < 1$ and the methods will yield the same results if RE=1.

3 Results and Discussion

3.1 Choosing the Correct Correlation Matrix for Firm Value Data

Based on the procedure outlined in Algorithm 1, the full GEE model is fit using five correlation matrices: Independence (IN), Exchangeable (EX), Order one Autoregressive (AR(1)), Toeplitz (TOEP) and Unstructured (UN). For each model, the EAIC and QIC values are obtained. The results are shown in Figures 1. and 2.

Figure 1 show that the model formulated with the AR(1) correlation matrix had a minimum EAIC value (967.7) hence was considerd the best for the data. The Unstructured matrix was the second least preferred by EAIC with the least preferred being the independence matrix. The second best preferred correlation matrix was Exchangeable.

Figure 2 show that, for the same data, QIC preferred the unstructured correlation matrix with a minimum value of 799.7 thus indicating QIC's preference for over-parameterized matrices. The Unstructured matrix preferred by QIC was the second

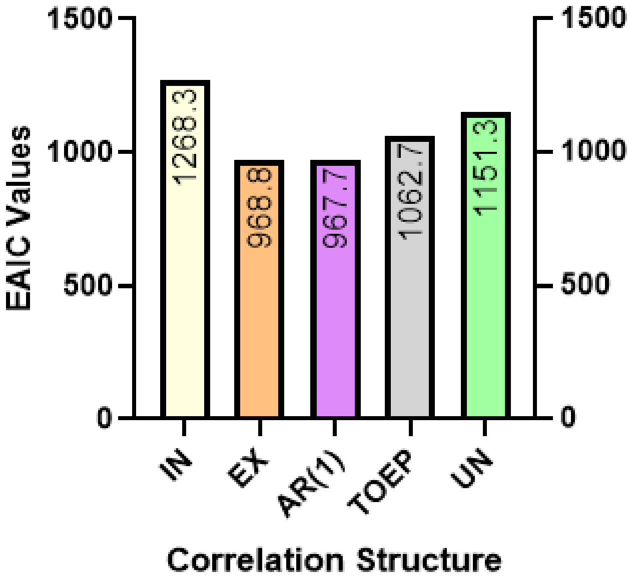

Figure 1: EAIC Values for Various Correlation Matrices

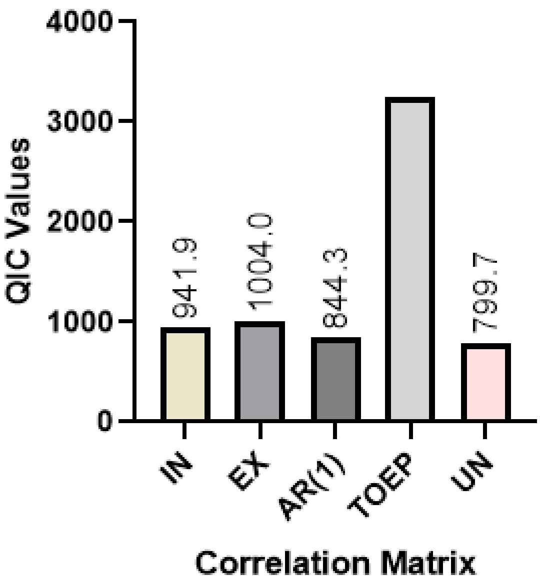

Figure 2: QIC Values for Various Correlation Matrices

least preferred by EAIC. AR(1) structure was ranked second by QIC with a value of 844.3. Based on Algorithm 1, it is concluded $R^s_{best} = AR(1)$. The results corroborate those of Chen and Lazar (2012) which suggested that EAIC had the tendency to choose a parsimonious structure more often and was more effective than the QIC in selecting the true correlation structure. It is therefore concluded that the AR(1) structure was a better fit for the data.

3.2 Selection of Regresssors Based on the AR(1) Matrix

Using $R^s_{best} = AR(1)$ and algorithm 2, the top ranked models based on their minimum QIC values and whose $\Delta_i \leq 2$ are in Table 1.

Table 1: Model Selection Table for QIC under AR(1) Matrix

M_K	Int	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	QIC	Δ_k
479	✓	✓	✓	✓	✓	✓	✓	✓				394.00	0.00
351	✓	✓	✓	✓	✓	✓	✓					394.71	0.71
349	✓	✓	✓		✓	✓		✓				395.26.	1.26
477	✓	✓	✓	✓	✓	✓		✓				395.38	1.38
223	✓	✓	✓		✓	✓	✓	✓				395.44	1.44
95	✓	✓	✓		✓	✓	✓					395.97	1.97

Table 1 shows that QIC selected the model with the regressors X_1, X_2, X_3, X_4, X_5, X_6 and X_7 as the best model for the shareholder value creation data. This implies that X_8, X_9 and X_{10} which represented Board size, accounting profitability and working capital policy respectively were dropped as drivers of value creation for firms in the NSE. To compliment the results in Table 1, variable relative importance (VRIMP) values were determined and are shown in Table 2.

Table 2: Variable Relarive importance measures for QIC under AR(1) Matrix

Covariate	X2	X1	X4	X7	X3	X6	X5	X9	X8	X10
VIMP	1.00	1.00	0.88	0.81	0.62	0.62	0.56	0.27	0.08	0.01

The results shows that X_1 and X_2 which represent growth rate of earnings and economic profitability respectively were the key drivers of value creation for firms in the NSE. Other important drivers in order of importance are X_4, X_7, X_3,X_6 and X_5 all with VRIMP values of greater than 0.5. Just like the results in Table 1, Board size, Accounting profitability and working capital policy had low VRIMP values hence considered not key drivers to value creation. However, it is worth noting that working capital ratio is a key component of the Altman's Z value that measures the financial health of the firms hence can be said to have an indirect effect on firm value. To ascertain the importance of selecting the true within subject correlation matrix in GEE modelling, the procedure in algorithm 2 was repeated using the unstructured correlation matrix that was chosen by QIC. Models whose $\Delta_i \leq 2$ are given in Table 3:

Table 3: Model Selection Table for QIC under AR(1) Matrix

M_K	Int	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	QIC	Δ_k
479	✓	✓	✓		✓	✓	✓	✓	✓			401.00	0.00
351	✓	✓	✓		✓	✓	✓		✓			402.49	0.49

The results indicate that the QIC-only benchmark method leads to the model with explanatory variables X_1, X_2, X_4, X_5, X_6, X_7 and X_8 being ranked as the top model for the data with a QIC value of 401 which was much greater than that of the models whose $\Delta_i \leq 2$ when the AR(1) structure was used. Notably, the top model included X_8 which was not in any of the top six models when the EAIC-QIC strategy was used, and that X_3 was excluded from the top model. Accounting profitability and working capital policy were both not included as drivers of value creation for firms in the NSE for the top two ranked models under this strategy. Likewise, under the QIC-only strategy, only two models had $\Delta_i \leq 2$.

The relative efficiency of estimators from the top-ranked models under the two approaches was compared using 5-fold cross-validation. The $MSE^{EAIC-QIC}$, $MSE^{QICOnly}$ and relative efficiency values for 10 iterations are given in Table 4

Table 4: Model Selection Table for QIC under AR(1) Matrix

Iteration	1	2	3	4	5	6	7	8	9	10
$MSE^{QICOnly}$	9.59	9.22	10.02	10.30	10.03	7.13	9.82	9.73	10.33	9.83
$MSE^{EAIC-QIC}$	7.57	7.94	1.02	1.02	1.01	5.22	8.03	7.79	1.01	8.60
RE	1.27	1.16	9.88	10.14	9.93	1.28	1.22	1.25	10.18	1.14

Table 4 shows that all the RE values were greater than one. This implies that the model selected under the proposed EAIC-QIC Hybrid methodology yielded more efficient estimates compared to the model selected under the QIC-only strategy. It is therefore inferred that the proposed Hybrid methodology greatly enhanced the GEE estimates' efficiency. The high mean squared error for the model under the QIC-only benchmark approach could be attributed to the use of the unstructured matrix which had more nuisance parameters to estimate thus costing efficiency (Chen and Lazar , 2012). The greater efficiency of the hybrid methodology can be attributed to the intracluster correlation structure selected by EAIC which if accurately modeled, enhances the efficiency of analysis using generalized estimation equations (Hin et al. , 2007). Based on the results in the selection table for the best model using the Hybrid methodology and Equation (16), the probability that a firm 'i' creates value for its shareholders in the t^{th} period will be:

$$\mu_{it} = \frac{Exp(\beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_4 X_{4it} + \beta_5 X_{5it} + \beta_6 X_{6it} + \beta_7 X_{7it})}{1 + Exp(\beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_4 X_{4it} + \beta_5 X_{5it} + \beta_6 X_{6it} + \beta_7 X_{7it})} \qquad (25)$$

3.3 Regularization of the best model Selected using the Hybrid Methodology

The best model selected using the hybrid methodology is subjected to SCAD regularization to weed out non-informative covariates. The regularization parameter λ was obtained using 5-fold cross-validation with 30 iterations. Findings are illustrated in Figure 3.

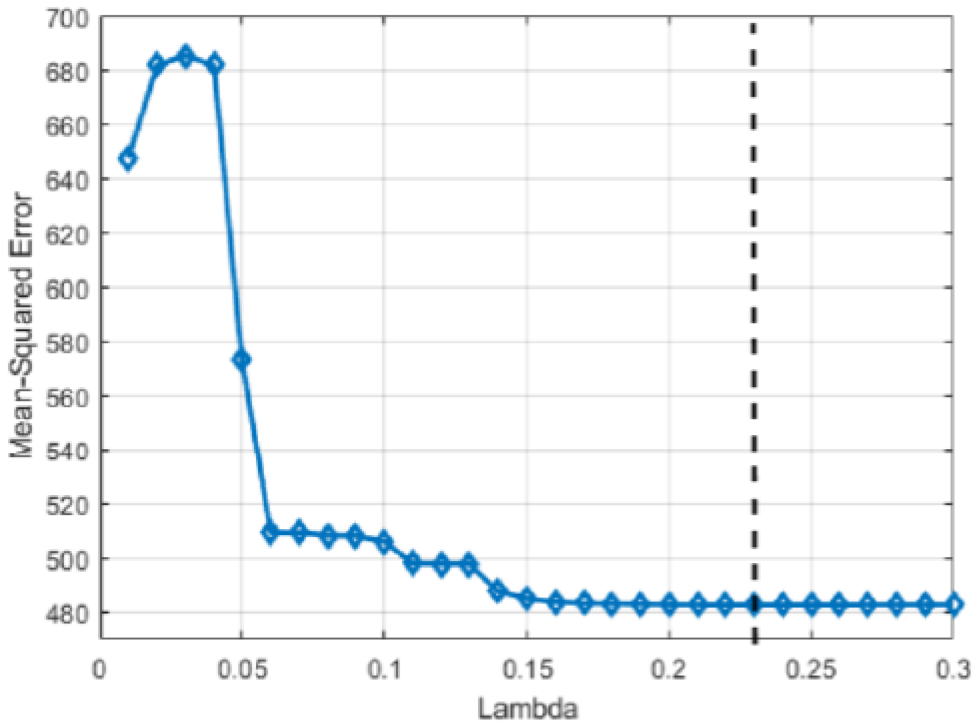

Figure 3: Optimum value of λ for SCAD Regularization

The results in Figure 3 show that $\lambda = 0.23$ was the optimum tuning parameter and was therefore applied in fitting the PGEE model. With the SCAD penalization, the regressors X_4, X_5 and X_6 which represent firm size, leverage and dividend policy are dropped from the model and only the regressors X_1, X_2, X_3 and X_7 which respectively represent economic profitability, growth rate of earnings, the interaction between the two and level of financial health were retained as the main predictors of firm value. The deletion of financial leverage from the final model supports MM capital structure irrelevance policy of 1963 by Modigliani and Miller which suggested that financing decision does not matter in value creation. The deletion of firm size from the SCAD regularized model confirms findings by Mule et al. (2015) who established that firm size had no significant effect on firm value. Also, the deletion of dividend policy meaning that it is not an important predictor of value creation corroborates findings by Hansda, et al. (2020) and Nguyen and Faff, (2007) but is contrary to findings by Agung et al. (2021a) and Bataha et al. (2023) who both established a positive significant effect. The deletion of dividend policy can however be justified since dividend pay-out ratio has a direct effect on growth in earnings (g) which also depend on Return on Equity (ROE) since $g = ROE \times \gamma$, where γ is the retention ratio such that $(1-\gamma)$ is the dividend pay-out ratio. The selected covariates, their coefficients, naive and standard errors are shown in Table 5.

Table 5: PGEE Regression Coefficients and Standard Errors

Criteria		Int	X1	X2	X3	X7
Penalized GEE	Estimate	-1.076^{***}	-0.3723^{***}	-0.6044^{***}	0.0183^{*}	0.0000143^{*}
	Naive SE	0.1752	0.0923	0.1459	0.0373	0.0001
	Robust SE	0.1731	0.0948	0.1509	0.0078	0.00004
UnPenalized GEE	Estimate	-2.781^{**}	0.0781^{**}	-7.406^{**}	0.0054	0.0184
	Naive SE	1.5040	0.0273	2.475	0.0048	0.0979
	Robust SE	1.0352	0.0311	2.7347	0.00247	0.0604
Signif. codes		$'{***}'$, 0.001	$'{**}'$, 0.01	$'{*}'$, 0.0		

The PGEE estimators indicate that the four regressors retained were all statistically significant at 5% level of significance with economic profitability and growth rate of earnings exhibiting a negative relationship with $\frac{M_v}{B_v}$ ratio. This meant that a firm that seeks to increase its economic profitability and growth in earnings will compromise value creation for their shareholders. On the other hand, financial distress had a positive effect on value creation. This means that firms that had smaller financial distress values meaning higher Altman's Z score values will have higher likelihoods of creating value for their shareholders compared to distressed firms. Based on the PGEE estimates, the model for the chances of a firm i creating value at time t takes the form;

$$\begin{aligned} \mu_{it} &= \frac{Exp(\beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_7 X_{7it})}{1 + Exp(\beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_7 X_{7it}} \\ &= \frac{Exp(-1.076 - 0.3723 X_{1it} - 0.6044 X_{2it} + 0.0183 X_{3it} + 0.0000143 X_{7it})}{1 + Exp(-1.076 - 0.3723 X_{1it} - 0.6044 X_{2it} + 0.0183 X_{3it} + 0.0000143 X_{7it}} \end{aligned} \tag{26}$$

Based on the results in Theorem 2.3, assuming that the regressors X_1, X_2, X_3 and X_7 are not at play, the probability of a firm 'i' in the NSE creating value for its shareholders is;

$$P_0 = \frac{Exp(-1.076)}{1 + Exp(-1.076)} = 0.2543 \tag{27}$$

which implies that the number of times the said firm is expected to create value in m years will be;

$$\begin{aligned}\lambda_i &= \frac{mP_0}{1-(1-P_0)^m} \\ &= \frac{0.2543m}{1-(1-0.2543)^m}\end{aligned}$$

Example 3.1. If m=3, $\lambda_i = \frac{0.2543\times 3}{1-(0.7457)^3} \approx 1$, if m=6, $\lambda_i = \frac{0.2543\times 6}{1-(0.7457)^6} \approx 2$, If m=10, $\lambda_i = \frac{0.2543\times 10}{1-(0.7457)^{10}} \approx 3$ e.t.c

The corresponding cumulative distribution function curve is shown in Figure 4

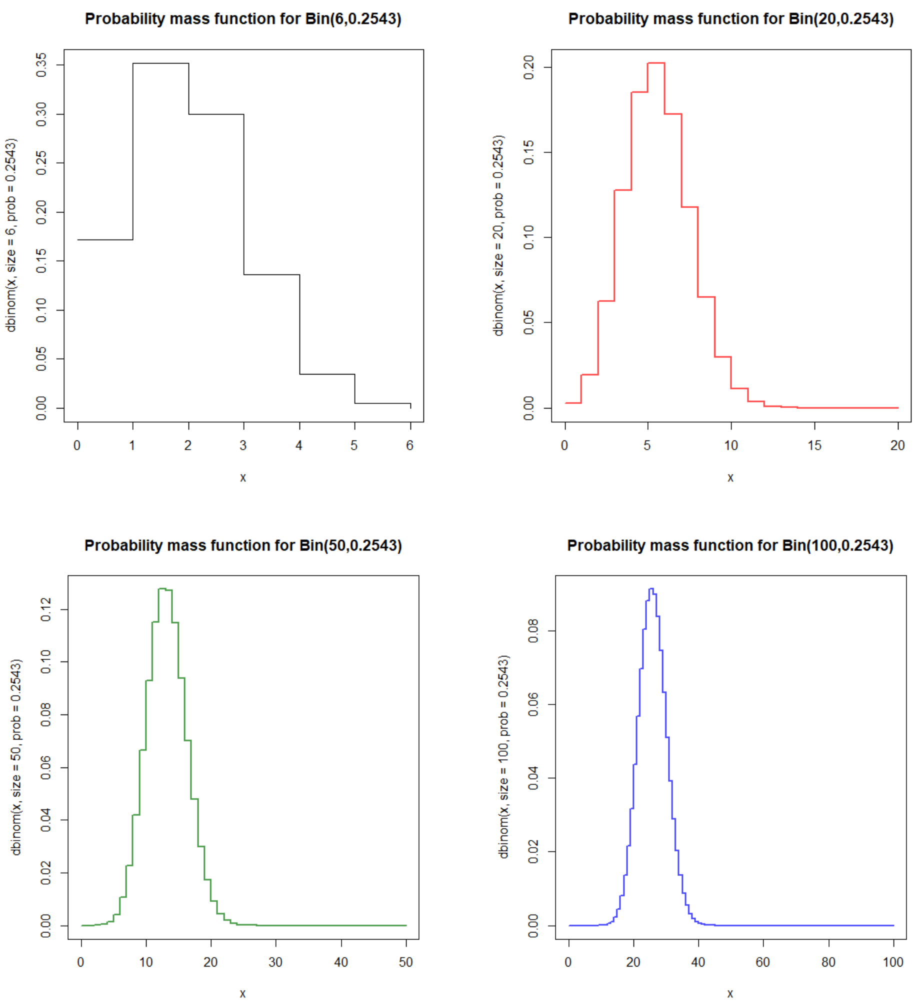

Figure 4: Binomial distribution for various 'm' values and P_0=0.2543

According to the plots in Figure 4, the graphs are positively skewed, and the skewness decreases as n (years considered) increases.

4 Conclusions

In this article, a Two-Step hybrid methodology that combines EAIC and QIC to select the true correlation matrix and set of regressors respectively was proposed. The proposed procedure was used to develop a GEE model for the drivers of firm value. Compared to the QIC-only benchmark method, the proposed method performed well in selecting a GEE model whose estimates were more efficient than those under the QIC-only approach. Also, the use of penalized GEE, ensured sparsity of the final model.

a In relation GEE modelling, it can be concluded that;

  i For efficiency improvement, it is pertinent that the correct correlation matrix be selected first based on the full model and then applied in the selection of the correct set of regressors. This will enhance efficiency of the resultant model.

  ii Hybridization of the model selection procedure in GEE modelling improves the GEE estimators' efficiency. This is true since the combination of EAIC and QIC yielded a versatile tool that ably overcame the established weaknesses of QIC in selecting the true correlation matrix, but capitalized on its strength of being impressive in variable selection.

  iii Sparsity in GEE model selection can only be achieved if techniques that incorporate penalization such as PGEE are used in model fitting. This ensures that redundant variables that only add noise to the final model are discarded.

b In relation to drivers of shareholder value creation, it can be concluded that;

  i The Gordon-Constant growth model continues to be vital in evaluating value creation by listed firms

  ii Dividend policy has no direct effect on shareholder value creation but rather an indirect one through the growth rate of earnings (g), which is a function of the dividend pay-out ratio.

  iii Since working capital is a key component in the computation of the Altman's Z score used o determine the financial health status of the firms, it is concluded that working capital policy will have an indirect effect on shareholder value creation. This follows from the results that financial health has a significant effect on value creation.

Acknowledgement

I would like to thank the anonymous editor and the reviewers for their constructive comments that helped improve this paper.

References

Agung, G. and Hasnawati, S. and Huzaimah, R. F. (2021a). The Effect of Investment Decision, Financing Decision, Dividend Policy on Firm Value. *Jurnal Bisnis dan Manajemen (JBM)*, 17(1): 1-12.

Bataha, I. E., Ilat, V. and Kalalo, M. Y. B. (2023) The Effect of Dividend Policy and Market to Book Value of Equity on Firm Value in Energy Sector Companies Listed on Indonesia Stock Exchange. *Jurnal Ekonomi dan Bisnis Digital (MINISTAL)*, 2(1):29–42

Burnham, K. and Anderson, D. (2002). Model Selection and MultimodelInference: a practical information-theoretic approach 2nd edition. *Springer-Verlag, New York.*

Carey, V. J. and Wang, Y. G. (2011). Working Covariance Model Selection for GEE. *Journal of Statistics and Medicine*, 70(26): 3117-3124.

Chen, J. and Lazar, N. A. (2012). Selection of Working correlation structure in Generalized Estimating Equations via Empirical Likelihood. *Journal of computational and graphical Statistics*, 21(1):18–41.

Chen, L. J. and Chen S. Y. (2011). The influence of profitability on firm value with capital structure as the mediator and firm size and industry as moderators. *Investment Management and Financial Innovations*, (8, Iss. 1): 121–129.

Cui, J. and Qian, G. (2007). Selection of Working Correlation Structure and Best Model in GEE Analyses of Longitudinal Data. *Communications in Statistics—Simulation and Computation*, 36(5): 987–996.

Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of American Statistical Association*, 96(456:) 1348–1360.

Gordon, M. J. (1959). Dividends, Earnings, and Stock Prices *The Review of Economics and Statistics*, Vol. 41(2): 99-105.

Hansda, S., Sinha, A., and Bandopadhyay, K. (2020). Impact of Dividend Policy on Firm Value with Special Reference to Financial Crisis *SIT Journal of Management*, 10(2): 158-175

Hin, L.-Y. and Wang, Y.-G. (2009). Working Correlation Structure identification in generalized estimating equations. *Statistics in Medicine*, 28(4): 642-658.

Hin, L.-Y. Carey, V. J. and Wang, Y.-G. (2007). Criteria for Working Correlation Structure Selection in GEE. Assessment Via Simulation. *The American Statistician*, 61(4): 360-364.

Honjo, Y. and Harada, N. (2006). SME Policy, financial structure and firm growth. Evidence from Japan. *Small Business Economies*, 27(4): 298–300.

Liang, K.-Y and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*,73(1):13-22.

Mule, K. R., Mukras M. S., and Nzioka, O. M. (2015). Corporate Size, profitability and market value: An econometric panel analysis of listed firms in Kenya. *European Scientific Journal*, 11(13):376–396.

Nguyen, H. and Faff, R. (2007). Impact of Board size and oard diversity on firm value: AUstralian Evidence. *Corporate Ownership and Control*, 4(2):24–32.

Nyabwanga, R. N., Onyango, F., and Otumba, E. O. (2019a). Consistyency Inference Property of QIC in Selecting the True Working Correlation Structure for Generalized Estimating Equations. *American Journal of Theoretical and Applied Statistics*, 8:74-84.

Nyabwanga, R. N., Onyango F. and Otumba E. O.(2019b). Inference properies of QIC in the selection of Covariates fo Generalized Estimating Equations. *Internartional Mathematical Forum*, 14(4):149–157

Olaniran, O. R. and Abdullah, M. A. A. (2019). Bayesian Variable Selection for Multi-class classification using Bootstrap Prior Technique. *Australian Journal of Statistics*, 48(2):63–72.

Pan, W. (2001). Akaike Information Criteria in generalized estimating equations. *Biometrics*, 57(1): 120–125.

Pandey, I. M. (2005). What drives Shareholder Value. *Indian Institute of Management, Ahmedabab, India Working Paper 2005-09-04*.

Wang, L., Zhou, J., and Qu, A. (2012). Penalized Generalized Estimating Equations for high dimensional longitudinal data analysis. *Biometrics*, 68(2):353–360.