© AUG 2022 | IRE Journals | Volume 6 Issue 2 | ISSN: 2456-8880

Application of Heckman-Conway-Maxwell-Poisson Model for Analysing Corruption

LEONARD KING'ORA THUO[1], DR. JOSEPH OUNO OMONDI[2], DR. BONIFACE KWACH[3]
[1, 3] *Kibabii University-Kenya*
[2] *Masai Mara University-Kenya*

Abstract- The main objective of this paper was to illustrate application of Heckman-Conway-Maxwell-Poisson 'HeckCOMPoisson' model for analyzing corruption. HeckCOMPoisson model is an integrated model from Heckman and Conway-Maxwell-Poisson model. HeckCOMPoisson was developed to handle count model with selection. The specific objective of this paper was to test the fitness of the integrated HeckCOMPoisson model in predicting number of people paying bribes. Corruption is among the most serious problems of the societies because it destroys the democratic progress by creating unemployment, increasing poverty level among other negative effects. To achieve the objective of this paper, summary tables were created, graphs were drawn and the results discussed extensively. This paper used the corruption data of 2019 from Ethics and Anti-Corruption Commission (EACC) of Kenya. The results of this study showed that HeckCOMPoisson perform perfectly in terms of Goodness-of-Fit (GOF) and in prediction of count data with selection and handling under-dispersion and over-dispersion experienced in count data. The results of this study proved that HeckCOMPoisson distribution is flexible and gives robust models for dispersed counts. By parametrizing HeckCOMPoisson distributions via the mean, variance and model prediction, this study placed HeckCOMPoisson models on the same level of interpretability and parsimony as other popular, competing count models.

Indexed Terms- Corruption, HeckCOMPoisson

I. INTRODUCTION

Corruption is a complex phenomenon from its numerous definition to its different forms and how it is practiced. Monte and Erasmo [11] used Ordinary Least Square (OLS) to analyze corruption where GDP was their dependent variable and a number of respondents who participated in corruption and corruption perception index were taken as explanatory variables. From the results of their model, they concluded that corruption encourages crime. Though these results seemed to concur with the results of other research studies, their model failed to account for the extreme large values of the standard deviations which were larger than the means of the study variables. This indicated that their data was more spread out. When there is a lot of variance in the observed data around the mean, there is a need to use a count model such as COMPoisson model that takes into account the dispersion of the data.

Using a Multi-channel queuing model, You and Khagram [17] argued that corruption leads to a negative effect on employment level, domestic investment and government spending. This model analysed all the variables in a single model without considering that some variables were subsets of the main variables and therefore they needed to be analyzed in a step model. In addition, this model treated corruption as a single variable but not a series of factors that would influence respondents' decision to participate in corruption which consequently affect employment level in the long run. Failing to put the variables in two sets creates a sample selection biasness since the variables are randomly included in the model. This limitation makes the results of queue model not adequately reliable.

Tirole [14] investigated the relationship between corruption and various crime series of Japan, Australia and South-Korea (that is, Asia pacific countries), using Auto Regressive Distributed Lag (ARDL) model to find whether there was a long-run equilibrium between crime and corruption. Tirole [14] found out that there was a strong positive relationship between crime and corruption. However, this time-series model helped in graphically displaying the patterns and trends of the data but did not give the probability of a respondent being involved in corruption. Additionally, the ARDL in their results had difficulties in successfully identifying the correct association between the variables in the data which contained a unit root (that is, a difference stationary process) hence leading to the issues of spurious correlation arising. A time series with a unit root shows a pattern that is unpredictable. This gap in ARDL can be addressed by a model such as HeckCOMPoisson model since HeckCOMPoisson has predictive margin plot which has systematic pattern that is predictable since it have predicted probabilities.

Adebayo [2] in his attempts to establish the nexus between unemployment and corruption between 2000 and 2009 in Nigeria using simultaneous equations model approach showed that there was a negative effect of corruption on unemployment. This model predicted a reduction in the rate of employment and economic growth with increase in corruption. However, this model only examined the impact of corruption on other factors in different simultaneous equations, ignoring the potential impact these factors would have if they were to be put in one equation to examine the effects they would have on corruption which was the dependent variable. Failing to examine the effects of these variables in a single equation injected a priori bias into the model. Hence, these models failed to quantify the potential tradeoffs between corruption and the study variables.

II. PROBLEM STATEMENT

One of the unsolved problems in analysis of corruption, is lack of a count-decision model that has minimised risk to produce biased and inferior parameter estimates. Corruption studies suffer from the bias of focusing in developing countries due to lack of an integrated approach, Andris [3]. According to United Nations Convention against Corruption [15], they reported that there is need to have measurements methods on corruption that can be used to identify trends and illustrate the scale and scope of particular types of corruption. From the research of Mungiu and Heywood [12], they concluded that progress in corruption research and anticorruption policies has been hampered by the lack of adequate measurement method. Additionally, according to Kannan [19], corruption will remain a challenge to address in future, if corruption measurements do not genuinely account the methodological realities. Previous models had the limitations of unaccounted dispersion, endogeneity bias, spurious correlation, unpredictable patterns among other issues. For these reasons, there was need to develop a model that can curb these gaps to attain a model with high accuracy in parameter predictions.

III. LITERATURE REVIEW

Decision models and count models are important tools in understanding factors influencing citizens to participate in corruption. There are only a few decision models applied in the studies on corruption. The study by Abdulrahman [1] proposed and analysed a deterministic model for corruption in a population. The researcher computed the basic reproduction number (BRN), corruption-free equilibrium point, and endemic equilibrium point. Numerical simulations were carried out and it was revealed that corruption can only be reduced to a manageable level but not totally removed.

Nathan and Jackob [13] developed an epidemiological compartment model for corruption in Kenya, mainly considering those who take advantage of office holders and political office holders. In their paper, an optimal control theory was used to study the effectiveness of all possible combinations of two corruption preventive measures, namely, (i) campaigning about corruption through media and advertisement and (ii) exposing the corrupted individuals to jail and giving punishment. The researchers concluded that jailing the corrupted individual was more effective.

Using simple Pearson's correlation coefficients, Huther and Shah [8] found that decentralization lowered corruption. However, these unconditional correlations did not control for other variables that happened to be associated with decentralization. Fisman and Gatti [7] also found that decentralization decreased corruption using OLS and a country's legal origin as an instrument. However, La Porta et al. [10] suggested that a country's legal origin affects a country's level of corruption directly, which was an issue [7] acknowledged undermines the validity of their instrument.

The study by Waykar [16] developed the differential equation-based models representing either growth or decay laws for corruption. The differential equation-based models were framed to measure the level of corruption. These corruption models showed that the complete prevention of corruption is possible if the ratio between rate of dismissal and rate of corruption is equal to one.

Game theory provides powerful tools for situations when some bargaining between economic agents is involved. For example, it can be assumed that there is recursive "Nash bargaining" between the bribe-taker and bribe-giver. Basu et al [5] made a game-theoretic computation of an equilibrium between the bribe-taker and bribe-giver. The author used a cross section study, which found a negative relation between corruption and wages, implying a need to raise wages of bureaucrats.

Previous studies on corruption did not analyse corruption in a two-step using a single model but instead they analysed all the variables (such as, education level, those who participated in) in a single model without considering that some variables were subsets of the main variables and therefore they needed to be analyzed in a step model. Since previous models did not use a step model, they failed to quantify the potential tradeoffs between corruption and the study variables. Further, the empirical approach to employ multivariate techniques that regress performance outcome variables on discrete measures of organizational choices potentially suffer from selection based endogeneity bias. Selection-effects represent an internal validity threat that can lead to biased parameters estimates hence rendering erroneous empirical results and incorrect conclusions. This study developed an integrated 'HeckCOMPoisson' model that could model all forms of corruption without endogeneity selection bias.

Artjoms and Timothy [4] in their study of bribing behavior and sample selection, they paid attention to the issue of respondents' non-random selection into contact with public officials, which may result in biased estimates. Their results using Heckman probit model results suggested that the elderly were less likely to bribe public officials, while linguistic minorities, people with higher perceived relative income and those with lower trust in public institutions were more likely to bribe. Additionally, their results also showed that not accounting for sample selection effects produces an upward bias in estimated coefficients.

Andrzej and Lukasz [3] in their research on, who suffers and how much from corruption, they analyzed two variables related to corruption: the perception of corruption as an obstacle to doing business and actual bribe tax payments. To test the robustness of their results, they modeled a two-step decision using the Heckman self-selection model. Controlling for other factors, both sets of their empirical results showed that the extent of corruption is related to the time spent dealing with regulations and inspections. They argued that firms which spend more time dealing with administrative procedures have a greater perception of corruption and are forced to make significantly higher bribe payments.

IV. METHODOLOGY

4.1 Integrated HeckCOMPoisson Model

Unlike the usual Heckman model, integrated HeckCOMPoisson is an appropriate estimating model that fits a COMPoisson regression model into the Heckman model and hence eliminates endogenous sample selection bias. Heckman model is not appropriate for count outcomes because its linear model for the outcome often produces negative predicted values and does not restrict the predicted values to integers, but the integrated HeckCOMPoisson solves this problem. Integrated HeckCOMPoisson is an integrated model for estimating parameters of a count-data model with endogenous sample selection.

In addition, drawing the correct inference about selection effects depends on collinearity between the inverse Mills ratio and the other predictors in the second stage equation. When the error terms from the selection and the outcome equations are correlated (that is, $\rho \neq 0$), the standard probit techniques yield biased results, Breslow [6]. However, there are few methods to correct the standard errors in the second stage. The two documented methods are manual matrix manipulation and automatic correction using statistical packages. Endogeneity issue is the major limitation in Heckmam model. Endogeneity technically rises from correlation between the variables in the probit equation and select equation. Endogeneity biases parameter estimates. This research study added COMPoisson model in the second stage as an endogeneity correction approach. COMPoisson introduced a parameter denoted by ν that governed the rate of decay of successive ratios of probabilities and was not correlated with variables in the probit model. That is, Corr[probit(x),COMP(x')] = 0

4.1.1 Integrated HeckCOMPoisson Stage One Probit Function

The integrated HeckCOMPoisson model has two equations, one equation for the count outcome, *y*, and another equation for a binary selection indicator, *z*. The indicator z takes values of 0 or 1. The outcome *y (number of times a citizen* is observed only if z = 1, that is, a citizen adopts civic education on corruption. In summary, the count variable Y_i is assumed to have a COMPoisson distribution, conditional on the covariates X_i, with conditional mean given by the equation below. In this study, a corruption case was counted if a citizen gave or took a bribe. Other covariates that will be used to determine number of corruption cases are; adoption of civic education of corruption, amount of bribe taken or given, effectiveness of EACC materials among other covariates.

$$E(Y_i \mid X_i, \varepsilon_{1i}) = \lambda \frac{\partial \log Z(\lambda,\nu)}{\partial \lambda} \approx \lambda^{\frac{1}{\nu}} - \frac{\nu-1}{2\nu} + \varepsilon_{1i} \quad (1)$$

4.1.2 Integrated HeckCOMPoisson Stage Two Outcome Function

We only observe outcome Y_i for observation *i* if z_i, the selection outcome, which is the binary outcome from a latent-variable model with covariates Θ_i is equal to 1, that is;

$$z_i = \begin{cases} 1, & if\ \Theta_i\gamma + \varepsilon_{2i} > 0 \\ 0, & elsewhere \end{cases} \quad (2)$$

Z_i - is a dummy variable taking value 1 if a citizen adopt civic education and public sensitizations measures to reduce corruption level in Kenya

Θ_i are variables hypothesized to affect adoption of civic education and public sensitizations measures to reduce corruption level (as gender, reporting of bribing cases, age of respondent, employment, effectiveness of EACC materials and form of corruption)

The error terms ε_{1i} from (1) and ε_{2i} from (2) have bivariate normal distribution with zero mean and covariance matrix, $\begin{bmatrix} \sigma^2 & \sigma\rho \\ \sigma\rho & 1 \end{bmatrix}$ where σ and ρ have their usual interpretation for the bivariate normal distribution. A nonzero ρ implies that the selected sample is not representative of the whole population and hence inference based on standard Poisson regression using the observed sample is incorrect.

To compute the matrix above, HeckCOMPoisson uses the truncated value of ρ. This enables the estimate of σ to be made consistent with the truncated estimate of ρ and therefore,

$$\sigma' = \boldsymbol{\beta}_i \dot{\rho} \tag{3}$$

Where $\boldsymbol{\beta}_i$ is the coefficient of the truncated rho.

Both the truncated ρ and the new estimate of σ' are used in all computations to estimate the two-step covariance matrix. The truncated rho lie in the range [-1;1]. If the two-step estimate for ρ is less than -1 then ρ is set to -1 and if the two-step estimate is greater than 1, ρ is set to 1.

In order to retain ρ within the valid limits described above and for numerical stability during optimization, the integrated HeckCOMPoisson estimates the inverse hyperbolic tangent of ρ using the equation below.

$$\text{atanh}\ \rho = \frac{1}{2} ln\left(\frac{1+\rho}{1-\rho}\right) \tag{4}$$

In Stata, the Integrated HeckCOMPoisson automatically computes the inverse hyperbolic tangent of ρ. Similarly, the Integrated HeckCOMPoisson does not directly estimate σ, for numerical stability it estimates $ln\ \sigma$. Estimation of ρ and σ in the forms atanh ρ and $ln\ \sigma$ extends the range of these parameters to infinity in both directions. Additionally, ρ and σ were used to compute λ which represented selection effect and was computed as

$$\lambda = \sigma\rho \tag{5}$$

The standard error of λ was computed using the delta method (that is, propagation of error method) as shown in (6) below.

$$Var(\lambda) \approx D\ var\ (\text{atanh}\ \rho\ ln\sigma)\ \check{D} \tag{6}$$

Where D is the Jacobian of λ with respect to atanh ρ and ln σ

4.2 Joint Likelihood Estimation of HeckCOMPoisson

In this study, various parameter of interest were estimated through the HeckCOMPoisson joint likelihood estimation equation. These parameter include; bribe amount a citizen gave or received, the leading form of corruption, the services that people paid most bribe to be assisted, awareness of EACC materials, effectiveness of EACC and effectiveness of media in fight against corruption. The joint log likelihood of HeckCOMPoisson was computed as shown by (7) below.

$$\ln L(\pi) = \sum_{i=1}^{N}[Z_i \times ln\{\Pr(Y_i, z_i = 1)|X_i, \Theta_i, \pi\} + (1 - Z_i) \times \ln\{\Pr(Z_i = 0|\Theta_i, \pi)\}] \tag{7}$$

Where;

Y_i - is a dummy variable taking value 1 if a citizen has bribed and 0 otherwise

Z_i - is a dummy variable taking value 1 if a citizen adopt civic education and public sensitizations measures to reduce corruption level in Kenya

X_i is the regressor variable and in this study, these are variables hypothesized to affect adoption of civic education and public sensitizations measures to reduce corruption level

Θ - is the standard normal cumulative distribution function,

The joint probability $\Pr(Y_i, z_i = 1|X_i, \Theta_i, \pi)$ was obtained by integrating the conditional probability $\Pr(Y_i, z_i = 1|X_i, \Theta_i, \pi, \varepsilon_1)$ over ε_1. This can be expressed as shown in the equation below.

$$\Pr(Y_i, z_i = 1)|X_i, \Theta_i, \pi) = \int_{-\infty}^{\infty} \Pr(Y_i|X_i, \varepsilon_1)\omega\left(\frac{\Theta_i\gamma + \frac{\rho}{\sigma\varepsilon_1}}{\sqrt{1-\rho^2}}\right)\emptyset\left(\frac{\varepsilon_1}{\sigma}\right)d\varepsilon_1 \tag{8}$$

Where $\emptyset(.)$ is the standard normal density function and ω(.) is the standard normal cumulative density function.

Similarly, $\Pr(Y_i, = 0|\ \Theta_i, \pi)$can be derived as shown below.

$$\Pr(Y_i, = 0|\ \Theta_i, \pi) = \int_{-\infty}^{\infty} \omega\left(\frac{\Theta_i\gamma + \frac{\rho}{\sigma\varepsilon_1}}{\sqrt{1-\rho^2}}\right)\emptyset\left(\frac{\varepsilon_1}{\sigma}\right)d\varepsilon_1 \tag{9}$$

The integrations in (8) and (9) have no closed form are approximated using Gauss-Hermite quadrature. The integrated HeckCOMPoisson supports the Huber White [49] estimator of the variance.

Computing the joint log likelihood of the integrated HeckCOMPoisson can be time consuming with large datasets. For instance in this study where variables had more than 5900 entries, it would take more than six hours to compute the output. Solving this, a command *two-step* was added in the algorithm. Additionally, to achieve more accurate results from log likelihood approximation, this model used 25 quadrature points.
The integrated HeckCOMPoisson model does not need to run a separate probit or logit for sample inclusion followed by a regression as it is the case in the ordinary Heckman model. The integrated HeckCOMPoisson is the appropriate model to analyse and predict a count dependent variable that is observed in the absence of selection.

4.3 Computation of the Incidence Rate Ratios
The term incidence rate refers to the rate at which a new event occurs over a specified period of time. This rate only uses new cases rather than previously diagnosed or reported ones. The incidence rate is a very important metric for tracking change in corruption within a specific population over time. In this study incidence rate is the number of new corruption cases in the year 2019 (the numerator) as a proportion of the number of people involved in corruption (the denominator).

The incidence rate measures how often corruption incidences are likely to occur over a particular period of time and hence this metric will enable leaders to take action to remedy policies, including better regulations to increase options available to curb increase in corruption level. It is worth noting that incidence is different from prevalence, which measures the total accumulation of cases while incidence measures the likelihood of occurrence during a specific time period. Prevalence, on the other hand, is a measure of the actual number of cases of a condition in a population at a certain point in time. Therefore, it is the total accumulation of incidences over a period of time.

The HeckCOMPoisson stage two outcome function (that is, equation 2) is an integral part of the model specifying sample-selection. The integrated HeckCOMPoisson uses jackknife method to generate standard error that are robust and allow for intragroup correlation (that is, clustering). The integrated HeckCOMPoisson will generate and transform estimated coefficients to incidence-rate ratios, (that is, e^{β_i}instead of β_i). This study views some parameters as incidence-rate ratios (IRRs), (that is, we hold the entire x's in the model constant except one). The IRR for a one-unit change will be computed as shown in the equation below

$$\frac{e^{\ln(E)+\beta_i(x_{i+1})+\cdots+\beta_k x_k+\varepsilon_1}}{e^{\ln(E)+\beta_i x_i+\cdots+\beta_k x_k+\varepsilon_1}} = e^{\beta_i} \qquad (10)$$

4.4 Model Validation

4.4.1 Marginal Effect
One of the most paramount issues in developing a count data model with selection, or any model, is the interpretation of the coefficients. Computing the marginal value of a particular variable provide valuable information about how the model coefficient related to that variable influence the expected mean value. In this study, the calculation of the HeckCOMPoisson marginal effect was more complicated than the usual Negative Binomial model. This is because the parameter μ for the HeckCOMPoisson model is a centering parameter but for the Negative Binomial it is the expected mean value.

The relative marginal effect of a particular variable or covariate X_i was estimated using the following expression by Cameron [11];

$$\frac{1}{E[Y|X_i]}\frac{\delta E[Y|X_i]}{\delta X_i}$$

This study adopted (11) below to estimate the relative marginal effect for HeckCOMPoisson.

$$E[Y] \approx \mu + \frac{1}{2v} - \frac{1}{2} \qquad (11)$$

Equation 11 was to estimate the relative marginal effect of the number of times a citizen gave a bribe based on influencing factors (independent variables such as gender, reason for bribing, service bribed for, reporting of bribing cases, age of respondent, employment, effectiveness of EACC materials and form of corruption)

V. MODEL RESULTS

5.1 Description of the 2019 Corruption Data

This subsection briefly analysed the 2019 corruption data in order to obtain its general distribution. The data had twenty seven variables and 5942 observations. This study utilized only key variables to test the fitness of HeckCOMPoisson.

5.1.1 Descriptive Summary of Corruption Data

The results of table 5.1 indicated that out of a sample of 5942 who participated in the 2019 survey, 4111 (69.2%) had given bribe. It is also evident that, number of times a person was involved in giving bribes reduced with increase in frequency. The possible reason for this is, citizens tend to give up losing their money on corruption if they do not get the services they were bribing for. Additionally, by the time a person offers a bribe the tenth time, one will have achieved what he or she want or will opt to look for other means for solving problems he or she had.

Table 5.1: Summary Statistics for the 2019 Corruption Dataset

No. of times bribed (last 3 years)	Frequency	Percentage
0	1831	30.8%
1	1404	23.6%
2	1155	19.4%
3	930	15.7%
4	295	5.0%
5	167	2.8%
6	77	1.3%
7	11	0.2%
8	38	0.6%
9	26	0.4%
10+	8	0.1%
Total	5942	100%

5.2 Results of the Integrated HeckCOMPoisson Regression model

The developed HeckCOMPoisson model was executed and the results to test its fitness are as shown in table 5.2 below. The dispersion parameter was v = 0.08091, which indicated over-dispersion in the data. It was evident that there was some dispersion in the data since out of 5942 number of observations, 4111 were censored and 1831 were not selected into the Heck COMPoisson model.

The integrated HeckCOMPoisson estimated the inverse hyperbolic tangent of ρ (that is, */athrho*) to be -3.330218 using (12) below.

$$\operatorname{atanh} \rho = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho}\right) = -3.330218 \qquad (12)$$

HeckCOMPoisson further estimated $ln\ \sigma$ (that is, natural logarithm of sigma) to be -0.0261. The product of these two parameters (that is, atanh $\rho \times ln\ \sigma$), were used to compute the HeckCOMPoisson selection effect which was 8.658%. The interpretation of this percentage is that, there was 8.658% level of biasness in HeckCOMPoisson model. The selection effect was the level of bias which was introduced into the model after including a subset of target group. The target group was those citizens who only participated in corruption. This selection effect was relatively low and hence the model was fit.

Table 5.2: HeckCOMPoisson Model with Output with endogenous selection

HeckCOMPoisson Regression (25 quadrature points) *Log likelihood = -8810.022*	*Number of observation = 5942* *Selected = 4111* *Nonselected = 1831* *Wald chi2 (98) = 70.33* *Prob > chi2 = 0.000*		
Number of times bribed	Estimate	Beta (β_n)	P.Value
Maximum bribe amount offered	4.148	1.32	0.020*
Reason for bribing	1.165	0.63	0.009**
Service bribed for	0.13	0.054	0.011*
Reporting of bribing cases	4.047	0.15	0.882
Sources of Corruption information	-2.005	-0.48	0.632
Employment	11.029	1.11	0.267
Effectiveness of EACC materials	-6.008	-0.77	0.094

Form of corruption	5.376	0.13	0.056
Constant	10.74	0.98	0.329
/athrho	*-3.3302*	-	-
/lnsigma	*-0.02617*	-	-
rho	*-0.99744*	-	-
sigma	*0.43288*	-	-
v	0.08091	-	-
MAD	*3.496*	-	-
MSPE	*11.985*	-	-

*Wald test of independence equation (rho = 0) and Where **and * indicate statistical significance at 1%, and 5% respectively.*

From the results in table 5.2, (1) was approximated as shown in (13) after substituting the values of each parameter.

$$\mathrm{E}(Y_i \mid X_i) = \lambda \frac{\partial \log Z(\lambda,\nu)}{\partial \lambda} \approx \lambda^{\frac{1}{\nu}} - \frac{\nu-1}{2\nu} = \lambda^{12.36} + 5.68 \qquad (13)$$

Where the rate λ was given by;

$$\log(\lambda) = 10.74 + 4.148\,(Maximum\ bribe\ amount\ offered) + 1.165\,\boldsymbol{I}(Reason\ for\ bribing) + 0.13\,\boldsymbol{I}(Service\ bribed\ for) + 4.047\,\boldsymbol{I}(Reporting\ of\ bribing\ cases) - 2.005\boldsymbol{I}(Sources\ of\ Corruption\ information) + 11.029\,\boldsymbol{I}(Employment) - 6.008\boldsymbol{I}(Effectiveness\ of\ EACC\ materials) + 5.376\,\boldsymbol{I}(Form\ of\ corruption) \qquad (14)$$

Where I(.) is an indicator function. The findings from (14) indicate that the variable employment had the highest effect on the number of times a citizen bribed and in the long run it affected the decision to participate in corruption. Ordinarily, jobless citizens will tend to bribe many times till the day they will secure a job. On the other hand, it was evident that EACC materials had the least influence on corruption level. This indicated that most citizens were not getting elaborate information on corruption.

5.2.1 Mean Absolute Deviance (MAD)

From table 5.2, the HeckCOMPoisson Mean Absolute Deviance (MAD) gave a measure of the average misprediction of the model. The mean absolute deviation was used to measure of variability of the model. It was computed using (15) below:

$$MAD = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| = 3.496 \qquad (15)$$

The mean absolute value of 3.496 implied that data point of the independent variables were not highly spread out from the mean number of times a citizen bribed which was the dependent variable.

5.2.2 Mean Squared Predictive Error (MSPE)

The HeckCOMPoisson Mean Squared Predictive Error (MSPE) was typically used to assess the error associated with validation. This value was computed using the (16) below and results are displayed in table 5.2.

$$MSPE = \frac{1}{n}\sum_{i=1}^{n} \{\hat{y}_i - y_i\}^2 = 11.985 \qquad (16)$$

Equation (16) indicated that the average distance between the observed values and the model predicted values was approximately 12 units. Since there is no standard value to measure if the Mean Squared Predictive Error is correct, it is only observed to be close to zero and therefore the value in this study was considered appropriate for the model.

5.3 Marginal Effect for the count model with selection
After running the HeckCOMPoisson model and fitting in the required variables, this study further tested the fitness of the model on how best it would estimate the effect of an increase of one variable on the number of times a citizen bribed. For example, we tested if increasing citizens' age by one year would result to an increase or decrease in the number of times a citizen bribed. In this case, the prior expectation was that, citizens with higher age would have bribed more times than the very young citizens. The marginal effects at the means were computed and displayed in table 5.3.

Table 5.3: Marginal Effect Output of Analysis of Factors Affecting Citizens' Participation in Corruption

Variable	Marginal effect	Z	P>[Z]

	(dy/dx)		
Gender	0.313	2.98	0.003**
Reason for bribing	0.048	2.41	0.016*
Service bribed for	0.297	3.04	0.002**
Reporting of bribing cases	0.015	1.78	0.075
Age of respondent	0.336	2.02	0.043*
Employment	0.084	1.37	0.172
Effectiveness of EACC materials	0.0005	0.11	0.913
Form of corruption	0.004	1.47	0.141

*Where **and * indicate statistical significance at 1%, and 5% respectively.*

Results of table 5.3 above indicated that the age of a citizen had the highest marginal effect on the number of times a citizen bribed. An increase of a citizen's age by one unit would result to an addition 33.6% number of times a citizen bribed. Gender of the respondents closely followed with marginal value of 31.3%. Apart from indicating close association between the two variables, it showed that the two variables highly influenced citizens' decision in participating in corruption. HeckCOMPoisson marginal effects were computed using the expression below.

$$\frac{1}{E[Y|X_i]}\frac{\delta E[Y|X_i]}{\delta X_i}$$

The results of table 5.2 were consistent with HeckCOMPoison coefficients in table 5.3 whether the variable *Effectiveness of EACC material* had the least marginal value and hence the least influence citizens' decision to participate in corruption.

5.3.1 Graphing the Margins of Gender and Age

Since Age had the highest marginal effects, this study plotted the graph to enhance visual analysis on how the two variables correlated. The results are displayed in figure 5.1 below. The Y-Axis of the graph gave the probability on a citizen giving bribe. It was evident that as age of a citizen increased, the likelihood of participating in corruption increased to a maximum value of 1(that is, the probability of a sure event). The graph depicted that male citizens were more likely to participate in Corruption than the female citizens. The predictive margins of gender were computed at 95% Confidence Intervals (CIs).

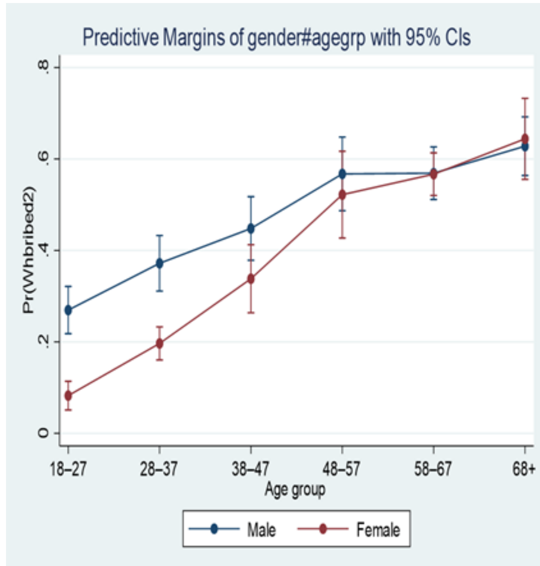

Fig 5.1: Predictive Margins of Age Against Gender

5.3 Computed HeckCOMPoisson Incidence Rate Ratios

A HeckCOMPoisson Incidence Rate Ratios in (10) was run with both categorical and continuous variables to determine the degree to which the independent variables influenced citizens' decision in participating in corruption. Results of the model were displayed in table 5.4 where the category with blanks was used as the base case.

Table 5.4: HeckCOMPoisson Incidence Rate Ratios

No. of times bribed		Coeff.	S.E.	Sig.	IRR
Employment	Employed	-	-	-	-
	Not employed	4.855	.357	.000**	128.407
Gender	Female	-	-	-	-
	Male	.241	.201	.230	1.272
Aware of EACC materials	Aware	-	-	-	-
	Not aware	1.669	.205	.000**	5.304
Form of Corruption	others	-	-	-	-
	favoritism	2.014	.843	.017*	7.496
	bribery	4.244	.461	.000**	69.662
	Fraud	2.229	.723	.000**	9.293
Reporting of bribing case	Reported	-	-	-	-
	Did not report	0.032	.161	.002**	.133
Reason for bribing	others	-	-	-	-
	To access services	2.987	.728	.000**	19.833
	to avoid paying full cost of services	3.212	.622	.000**	24.829
	To hasten the process	3.425	.644	.000**	30.710
Age-Group	18-27	-	-	-	-
	28-37	1.666	.664	.012*	5.289
	38-47	2.918	.708	.000**	18.511
	48-57	3.021	.755	.000**	20.517
	58-67	3.149	.632	.000**	23.309
	68+	3.526	.638	.000**	34.000

*25 quadrature points were used, Number of obs = 5942, Censored Obs = 4111, Uncensored Obs = 1831, Wald Chi2 (8) = 897.67, Prob > Chi2 =0.000, wald test of indep. Eqns (rho = 0) **and * indicate statistical significance at 1%, and 5% respectively*

The wald test Chi-square (at the bottom of the table) gave the validity of the HeckCOMPoisson model where the null hypothesis was that all the coefficients were zero. In this study the Wald Chi-square was 897.67 with p-value of 0.000 which showed correct model fit and fitted significantly better than a model with no predictors and therefore the null hypothesis was rejected at 5% level of significance.

As confirmed earlier from fig 5.1, there was strong association between Gender and participation in corruption. From table 5.4, Male citizens were 1.272 times more likely to participate in corruption than female citizens. Unemployed citizens were 128.407 times more likely to participate in corruption than the employed citizens and the IRR was statistically significant at 5% significance level. Initially, the variable employment was grouped into different categories (that is, Student, unemployed, self-employed-family, employed-private, employed-national government, employed-county government, employed-community, retired) but for this analysis it was coded as a binary variable where citizens were classified as either employed or not employed.

Citizens who had not read EACC materials were 5.304 time more likely to participate in corruption than those who had read EACC reports. This measured the level of civil education in fight against corruption. The corruption awareness IRR value of 5.304 was statistically significance with a p-value of 0.000.

Form of corruption which entailed, *bribery, abuse of office, intimidating behavior/abuse, conflict of interest, delays in service provision, embezzlement of public funds, favoritism, dealing with suspect property, lying/dishonesty, fraud/theft/graft, others* was recoded to 4 categories as shown in the table.

Bribery cases increased citizens' level to participation in corruption by 69.662 times more than other form of corruptions while fraud case increases by citizens participation in corruption by 7.496 times more compared to other form of corruptions.

From the 2019 EACC corruption survey data, citizens cited reason for bribing as*; it was the only way, to hasten the process, to avoid problems with authority, to avoid paying full cost of services, to access services and was expected.* This study coded the reasons into 4 categories (that is, *to access services, to hasten the process, to avoid paying full cost of services and others*). Setting *other reasons* as the reference group, hastening the process influenced citizens' involvement in corruption by 30.71 more times than other reasons.

As evident in figure 5.1 where increase in age increased probability of participating in corruption, the figure concurred with the gradual increase of the age group IRRs. Setting age group 18-27 as the reference group, age group 68 and above was 34.000 times more likely to participate in corruption than the age 18 – 17. These results confirmed that the older citizens had participated in corruption more times than the young citizens.

CONCLUSION

This study anticipate that the proposed HeckCOMPoisson model will form a building block for building more HeckCOMPoisson related models with either count regression problems or allow development of log linear models for counts. Additionally, the simple structure of the developed HeckCOMPoisson model is easy to understand on a conceptual level, and relatively easy to fit on a computational level and its computational times is significantly less compared to other count models. Finally, HeckCOMPoisson count regression models will forever remain to be a useful toolkit in any applied statistical field.

REFERENCES

[1] Abdulrahman, S. (2014). Stability analysis of the transmission dynamics and control of corruption. *Pacific Journal of Science and Technology,* vol. 15, no. 1, pp. 99–113.

[2] Adebayo, A. (2013). Youths' unemployment and Corruption in Nigeria: A nexus and implications for national development. Nigeria; Federal University.

[3] Andrzej, C. & Lukasz, G. (2020). Who suffers and how much from corruption? Evidence from level data. Journal of Economic Literature, 35(3), 2-23

[4] Artjoms, L. & Timothy, H. (2012). Bribing Behavior and Sample Selection: Evidence from Post-Socialist countries and Western Europe. Journal of Economic Literature. vol. 147(3), pages 377-393

[5] Basu, Kaushik, Sudipto Bhattacharya & Ajit M. (1992). Notes on Bribery and the Control of Corruption. *Journal of Public Economics* 48: 349-359.

[6] Breslow, N. (1990). Tests of hypotheses in over dispersed Poisson regression and other quasi-likelihood models. *J. Am. Statist. Ass.*, 85, 565–571

[7] Fisman, R. & Gatti, R. (2002). Decentralization and corruption: Evidence across countries. *Journal of Public Economics*, 83(3), 325-345.

[8] Huther, J. & Shah, A. (1998). Applying a simple measure of good governance to the debate on fiscal decentralization. World Bank,

[9] Kannan, P. (2021). Corruption Measurements: Caught Between Conceptualizing the Phenomenon and Promoting New Governance Agenda? Journal of Business Perspective. Volume: 26 issue: 1, page(s): 31-38

[10] La, P., Lopez, S., Shleifer, A., & Vishny, R. (1999). The quality of government. *Journal of Law, Economics and Organization*, 15(1), 222-279.

[11] Monte, Alfredo Del & Erasmo Papagni. (2001). Public Expenditure, Corruption, and Economic Growth: The Case of Italy. *European Journal of Political Economy,* 17, 1, 1-16.

[12] Mungiu, P. & Heywood, P. (Eds.). (2020). A research agenda for studies of corruption.Edward Elgar Publishing. Chapter 2.

[13] Nathan, O., & Jackob O. (2019). Stability analysis in a mathematical model of corruption in Kenya,*Asian Research Journal of Mathematics*, vol. 2019, pp. 1–15

[14] Tirole, J. (2006). 'A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality)', *Review of Economic Studies*, 63(1): 1-22.

[15] United Nations Convention against Corruption, (2015). National Anti-Corruption Strategies: A Practical Guide for Development and Implementation. New York

[16] Waykar, S. (2013). Mathematical modelling: a comparatively mathematical study model base between corruption and development, *IOSR Journal of Mathematics*, vol. 6, no. 2, pp. 54– 62, 2013.

[17] You, J., & Khagram, S. (2005). A Comparative Study of Inequality and Corruption. *American Sociological Review*, 70: pp.136-157.