Original Research Article

# An ecological adjusted random effect model for property crime in Windhoek, Namibia (2011-2016)

J. Amunyela[1], I. Neema[2], L. Pazvakawambwa[3*]

[1]Department of Mathematics & Statistics, Namibia University of Science and Technology, Windhoek, Namibia
[2]Namibia Statistics Agency, Windhoek, Namibia
[3]Department of Computing, Mathematical & Statistical Sciences, University of Namibia, Windhoek, Namibia

*Corresponding author: lpazvakawambwa@unam.na

ABSTRACT

Count data that are zero inflated are often analysed using Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) when observations are correlated in ways that require random effects. This study investigated ecological factors influencing the number of property crimes in Windhoek by using data obtained from the Windhoek police over the period of six consecutive years (2011 to 2016). The ecological concepts were measured at different levels of aggregation. Limited studies in Windhoek have considered analysing crime data on Generalized Linear Mixed Model via Template Model Builder (TMB) R-package. Crimes were counted with respect to Month, Season, Year, Location and Density. Property crime data contained more zeros than expected. When comparing models fitted, it was found that the Relative Risks (RR) were highly significant for models fitted via Negative Binomial distribution. By adopting a ZINB-GLMM, the study attempted to address the potential covariates for Property crimes. The study showed that most of the variation property crimes was due to locations. Crime was high during spring and winter time during the study period. The study further discovered that areas with high population densities had high crime intensity. Security patrols and surveillance should be stepped up in Windhoek in high density suburbs especially during winter and spring seasons.

## 1. Introduction

Crime, no matter how small, has an enormous negative impact on the people's lives. A rise in crime rate can negatively affect a country's economy as it discourages investors and pushed skilled workers elsewhere. Despite all the interventions and procedures put into place by policy makers, governments, and non-profit organisations to curb crime in Namibia, the crime rate continues to increase unabated, especially in the capital city, Windhoek.

An individual is less likely to be involved in criminal activities when there are substantial rewards and when he or she enjoys respect from the society in which they belong. In addition, if a young person is gainfully engaged, either in education or employment, he or she is less likely to turn to criminal activities (Khan, Ahmed, Nawaz, & Zaman, 2015). In most European and African countries population density was found to be appropriate in crime prediction (Hanley, Lewis, & Ribeiro, 2016). Tonry (2014) argued that the rate for property crimes rose recently in all wealthy Western countries. Property crimes in Western countries were declining as a result of improved security technologies in motor vehicles, residences, and retail stores (Tonry, 2014). This argument is also supported by the collaborative online database (NUMBEO, 2018). The NUMBEO crime index indicates that in the Caribbean alone, the average crime index is currently standing at 71% while the safety index is only at 29%.

Crime has negatively affected several African countries and this led to the identification of three main phases of crime due to tangible shifts in the prevailing social, political and economic environment (Shaw & Reitano, 2013). Crime and insecurity are major challenges in African countries, threats to national development and individual quality of life (Wambua, 2015). In addition to that, only 11 African countries ranked in the top hundred countries worldwide in terms of safety and security with Benin being the top-ranked African country at number 50 (Legatum Institute, 2014). Subsequently, 38% of Africans say they have felt

unsafe walking in the neighbourhoods. Study by Asongu and Kodila-Tedika (2016) found that in most African countries the wave of crime could be addressed if the fight against corruption is taken seriously by governments.

Drug trading, kidnapping, embezzlement, the large scale theft of minerals, or other plainly criminal activities exist in many parts of Africa (Ellis & Shaw, 2015). Palmary, Rauch, and Simpson (2014) contend that Johannesburg is believed to be the "crime capital" of South Africa. Besides that, a total of 267 arrests related to the illegal rhino horn trade were made in South Africa in 2012. This was linked to the arrest figures of 165 for 2010 and 232 for 2011 (Ayling, 2013). It was found that the illegal wildlife networks operating in South Africa and Namibia had no one distinct profile but both poachers and traffickers of rhino horn tended to be informal groups or predominantly individuals (Ayling, 2013).

Namibia is situated in Sub-Saharan Africa, a region that has one of the highest crime rates in the world (Neema & Böhning, 2012). According to the Overseas Security Advisory Council (Council, 2010), Namibians have regularly fallen victim to street crime.

Motor vehicle theft remains a major concern in Namibia. This type of crime usually involves smash-and-grab patterns and is sometimes associated with violence, especially when the occupants in the vehicle refuse to freely surrender their belongings to the perpetrators. Notably, Windhoek City Police (WCPOLS, 2006), observed that ATM card skimming, purse-snatching, vehicle breaks-ins and vehicle theft are among the most frequently recorded Property crime types in Namibia.

The introduction of Operation Kalahari and the installation of CCTV cameras was aimed at crime reduction and prevention in Windhoek. The Operation Kalahari and its predecessor Operation Horncranz involved Namibian Police force, Windhoek City Police and Namibian Defence Force working together to achieve a common goal. Personal robberies and residential break-ins and thefts remain prevalent in Namibia as well.

According to City of Windhoek (CoW), property crime's goal is to obtain money, property, or any other benefit. This may involve force, or the threat of force, in cases like robbery or extortion. This category includes, among other crimes, burglary, larceny, motor vehicle, arson, shoplifting and vandalism. High rates of Property crime were recorded recently at 74.06% in Windhoek (NUMBEO, 2018).

Scholarly opinion within the geography of crime and spatial criminology studies concurs that crime is highly concentrated in certain areas due to specific factors within those areas (Breetzke & Pearson, 2014; de Melo, Matias, & Andresen, 2015). Moreover, spatial patterns of these concentrations differ across crime types. Among youth, factors that influence crime are unemployment and lack of education (Dore, 2013), own house mortgage (Jones & Pridemore, 2012), crime-specific detection rate and prison population (Han, Bandyopadhyay, & Bhattacharya, 2013).

The basic model for count data like the number of property crimes is the Poisson model but most data does not satisfy the model assumptions. The study by Schielzeth and Nakagawa (2013), established that if the assumption made by random effects models are correct, then random effect would be the preferred choice because of its greater flexibility, generalizability, and its ability to model context, including variables that are only measured at the high level.

In contrary, findings by Brooks et al. (2017a) were that count data can be analysed using a generalized linear mixed model when observations are correlated in ways that require random effects. They further claimed that attempting to fit the GLMM models via a **glmmTMB** package with a log Normal-poisson model and covariate-dependent zero-inflation led to convergence failure, and hence they substituted with a similar model (a Negative-binomial model). A comparison of the results showed that the deterministic methods (gam, glmmTMB and inla) were all fast; gam was fastest, because gam fitted a simpler model. While on the other hand the stochastic methods (**MCMCglmm** and **brm**) were about an order of magnitude slower (Brooks et al., 2017a).

The Generalized Linear Mixed Model is of significance. GLMM include a random effect model which may be considered in this study based on its ability to estimate covariates both within and between clusters, its ability to partition variance at multiple levels, its ability to examine variation in effects across cluster and it is parsimonious. For completeness, the Poisson regression, Negative Binomial (NB) model and the Zero-Inflated Negative Binomial (ZINB) model were explored. These models were worth being explored since they are widely used on count data and produce reliable results (Harrison, 2014).

Bolker et al. (2009) acknowledge that count data with so many zero values cannot be made normal by transformation. Even when one succeeds to transform the data, this transformed data might violate some statistical assumptions or limit the scope of inference (one cannot extrapolate estimates of fixed effects to new groups). GLMMs combine the properties of two statistical frameworks that are widely used in ecology and evolution study. These are the linear mixed models (which incorporate random effects) and generalized linear models (which handle non-normal data by using link functions and exponential family e.g normal, Poisson or binomial distributions) (Bolker et al., 2009). Bolker et al. (2009) further concluded that GLMMs are the best tool for analysing non-normal data that involve random effects.

## 2. Zero Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM)

Count data has been analysed using generalized linear mixed models especially when observations are correlated in a way that require random effects. However, count data are often zero-inflated, containing more zeros than would be expected from the typical error distributions used in GLMMs (Brooks et al., 2017b). For example, crime counts may be exactly zero for some months based on effective policing but may vary according to the negative binomial distribution for an area with poor policing. In addition, the Zero-Inflation model which is part of ZINB-GLMM estimates the probability of an extra zero such that a positive contrast indicates a higher chance of absence (e.g. $\beta_i < 0$ means fewer absences in a variable $i$ unaffected by the fixed variable) while $\beta_i > 0$ means higher abundances in a variable $i$ unaffected by the fixed variable.

Bolker (2018); Brooks et al. (2017b) established a new R package, generalized linear mixed model Template Model Builder ($glmmTMB$), that increases the range of models that can easily be fitted to count data using maximum likelihood estimation. The interface is simply developed to be familiar to users of the $lme4$ R package, a widely used tool for fitting GLMMs. All one must do, in principle, is to specify a distribution, link function, and structure of the random effects (Bolker et al., 2009).

The Template Model Builder (TMB) was known for maximising speed and flexibility through utilising automatic differentiation to estimate model slopes and the Laplace approximation for handling random effects. The strength of the R package $glmmTMB$ lies on the number of benefits it poses. Among others, $glmmTMB$ is more flexible than other packages available for estimating zero-inflated models via maximum likelihood, and faster than packages that use Markov chain Monte Carlo sampling for estimation (Brooks et al., 2017b).

Furthermore, it is also rated high in terms of flexibility for zero inflated modelling than INLA even though speed comparisons vary with model and data structure. Study results by Bolker et al. (2009) were that repeated measurements on the same individual, the same location, or observations taken at the same point in time are often correlated and this correlation can be accounted for using random effects in a GLMM. The $MCMCglmm$ and $brms$ packages can fit zero-inflated GLMMs with predictors on zero-inflation, but they are relatively slow. Based on these limitations and challeges, the researcher preferred using the newly developed R package glmmTMB that can easily estimate zero-inflated GLMMs using maximum likelihood. The ability to fit these types of models quickly, using a single package made it easier to find the best model to explain patterns in the data. According to Berridge (2011), in GLMMs, the explanatory variables and the random effects (for a two level model, $x_{ij}, z_j$ and $\mu_{0j}$) affect the response (for a two-level model, $y_{ij}$) via the linear predictor ($\theta_{ij}$), where:

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \mu_{0j}. \quad (2.1)$$

Let $i = 1, 2, \ldots, m$ denote the index of the observational unit while $j = 1,2,3,\ldots,m_i$ denotes the index for the within observation in this unit. In this context $\theta = [\beta^T, \sigma^T]^T (d \times 1)$ denotes the vector of model parameters, where $\beta_{(p\times1)}$ represents the parameters for the fixed effect, and $\sigma_{(s\times1)}$ includes the parameters for the random effects while $d = p + s$. The observed outcome $y_{ij}$ is assumed to be independently drawn from exponential family of distribution when conditioned on a vector $X_{ij}(p \times 1)$ and the random effects vector $\gamma_i(q \times 1) \sim N_q(0, \Delta_\sigma)$. In this case $\Delta_\sigma$ is a positive-definite symmetric covariance matrix. For simplicity, the study considered reparametrization $\gamma_i = D_\sigma u_i$ resulting from the Cholesky decomposition of $\Delta_\sigma = D_\sigma D^T{}_\sigma$ where $u_i$ denotes the multivariate standard normal vectors. The GLMM is of the form:

$$g(u_{ij}) = \eta_{ij} = X^T{}_{ij}\beta + Z^T{}_{ij}D_\sigma u_i. \quad (2.2)$$

where $u_{ij}$ denotes the conditional expectation of the outcome, $Z_{ij}(q \times 1)$ a design vector for the random effects and $\eta_{ij}$ the linear predictor. Furthermore, $g()$ represents a link function which maps the linear predictor and the conditional expectation of the outcome (Flores-Agreda & Cantoni, 2019).

The GLMM is obtained by specifying some function of the response ($y_{ij}$) conditional on the linear predictor and other parameters, i.e.

$$f(y_{ij}|\theta_{ij}, \emptyset) = \exp\left\{\frac{[y_{ij}\theta_{ij} - b(\theta_{ij})]}{\emptyset} + c(y_{ij}, \emptyset)\right\}, \quad (2.3)$$

where $\emptyset$ is the scale parameter, $f()$ denotes the Probability Density Function (PDF), $b(\theta_{ij})$ is a function that gives the conditional mean ($\mu_{ij}$) and variance of $y_{ij}$, namely:

$$E[y_{ij}|\theta_{ij}, \emptyset] = \mu_{ij} = b'(\theta_{ij}), \quad (2.4)$$
$$Var[y_{ij}|\theta_{ij}, \emptyset] = \emptyset b''(\theta_{ij}), \quad (2.5)$$

while $c()$ is a funtion that is automatically determined once the other functions have been chosen (or simply denotes a specific function), so that the entire distribution is normalized.

In generalized linear mixed models, the mean and variance are related so that:

$$Var[y_{ij}|\theta_{ij}, \emptyset] = \emptyset b''(b'^{-1}(\theta_{ij})) = \emptyset V[\mu_{ij}]. \quad (2.6)$$

where $V[\mu_{ij}]$ is refered to as the variance function, $b'^{-1}(\theta_{ij})$ is a link function which expresses $\theta_{ij}$ as a function of $\mu_{ij}$, and $b'(\theta_{ij})$ is the inverse link function. The functions $b(\theta_{ij})$ and $c(y_{ij}, \emptyset)$ differ for different GLMMs. The distribution that works well in modelling the ZINB-GLMM is **nbinom2** by Magnusson et al., (2017), which returns an overdispersion parameter.

The expressions of the marginal PDF are obtained after integrating the random effects from the joint distribution $[y_{ij}, u_i^T]^T$.

Since the study data were counts which were zero-inflated, the Poisson, Negative Binomial, Zero-Inflated Poisson, and Zero-Inflated Negative Binomial models were fitted as alternatives to the Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM). The ZINB-GLMM could be more reasonable for this study because of its ability to handle multiple random effects components together with the zero-inflation and dispersion components. ZINB-GLMM performed well among other models for non-normal count data involving non-structural zeros due to its greater flexibility, generalizability, and its ability to model context, including variables that are only measured at high level.

The model selection was done using AIC and Residual plots. The AIC is a popular method for comparing the adequency of multiple, possibly non-nested models. The current practice is to accept a model with a small AIC value (Wagenmakers, 2004). The equation of the AIC is:

$$AIC = -2LL + 2r \qquad (2.6)$$

where $LL$ is a log likelihood value, $r$ indicates number of parameter and $n$ is a sample size (Posada & Buckley, 2004). In assessing the best fit model, the study further used the R package DHARMa by Walker (2018). DHARMa stands for "Diagnostics for HierArchical Regression Models". DHARMa includes support from glmmTMB and it is suitable for testing whether a GLMM is in harmony with the data (Dunn & Smyth, 1996). However, there are still a few limitations such as misspecifications in GLMMs which cannot be reliably diagnosed with standard residual plots.The expected distribution of the data changes with the fitted values and that makes GLMM residual harder to interpret. The current standard practice is to eye ball the residual plots for major misspecifications, potentially have a look at the random effect distribution and then run the test for overdispersion, but this approach still possesses a number of problems. The scaled (quantile) residuals are computed with the ***simulateResiduals( )*** function in R. The default number of simulation (n =250) was considered to be a reasonable deal between computation time and precision. What the function does is to create n-new synthetic datasets by simulating from the fitted model, compute the cummulative distribution of simulated values for each observed value and then return the scaled value that corresponds to the observed values (Walker, 2018).

The main objective of this research was to investigate ecological factors influencing property crime in Windhoek, to provide information and insight leading to better and improved prevention strategies of this category of crime. The specific objectives were to assess the ecological characteristics of property crime in Windhoek with a view to create understanding of its root cause; evaluate property crime in Windhoek and determine locations with high crime or crime hot spots for possible interventions; determine the season of the year in which property crime happen more often; and to model factors influencing property crimes in Windhoek and asses their impact.

## 3. Methods

A quantitative design was adopted in this study based on secondary data on daily reported Property crimes obtained from the Windhoek City Police department (2011 to 2016). The City Police Chief authorised the researcher to use the data. The daily crime data were recorded in the pocketbooks by the City police officers who attended to crime scenes.

After the pocketbooks were fully completed, they were submitted to the immediate supervisors or to the City Police Statistics Department for recording. In this study, the response variable was the Number of Property crimes $(Y_i)$. The independent variables $X_i$, for $i =1, 2….5$, refer to the *Month*, *Year*, *Location*, *Season* and *Density*. The variable *Month* represented a specific month from January to December in which crimes were committed in a specific year within a location.

The variable *Location* represented fifty-nine Windhoek geographical residential areas namely, Academia, Babylon, Brakewater, Cimbebasia, Damara Location, Dolam, Donkerhoek, Dorado Park, Dorado Valley, Eehambo NdaNehale, Eros, Eros Airport, Freedom Square, Freedomland, Gemeente, Golgota, Goreangab, Green-Well Matongo, Grysblock, Hakahana, Havana, Herero Location, Hockland Park, Independence Arena, Khomasdal, Kilimajaro, Kleine Kuppe, Klein Windhoek, Lafrenz, Ludwigsdorf, Marua Mall, Malaka Draai, Maroela, Mix, Mukwanangombe, Mukwanekamba, Northern Industrial, Nubuamis, Okahandja Park, Okuryangava, Olympia, Ombili, One-Nation, Ongulumbashe, Oshitenda, Otjomuise, Pionierspark, Properita, Rocky Crest, Shandumbala, Single Quarter, Southern Industrial, Soweto, Suiderhof, Vambo Location, Wanahenda, Windhoek Central, Windhoek North, and Widhoek West.

Prior to fitting the models for this study, a new variable S*eason,* which represents Summer (January-March), Autumn (April-June), Winter (July-September), and Spring (October-December), was created (Kemper & Roux, 2005).

The variable *Density* was obtained as the population per kilometer square of the area. The density was then scaled per 10 000 people. A good rule of thumb is that input variables should be small values, probably in the range of 0-1 or standardized with a zero mean and a standard deviation of one. The dataset was cleaned and re-coded prior to the data analysis.

To obtain the first overview of the dependent variable (Number of Property crimes), a histogram, boxplot and the normal Q-Q plot of the observed count frequencies were presented. Multiple plots of the Number of Property crimes were displayed to check the crime pattern on a yearly, seasonal and monthly basis via boxplots.The study further assessed the patterns of these crimes across all the fifty-nine locations.

## 4. Results

Cross tabulation were computed using the Statistical Package for Social Sciences (SPSS) version 25 (Green & Salkind, 2016). This was done to obtain the final counts of reported cases across the Property crime. Table 1 below shows a count summary of seasonal reported cases under Property crime, reported from 2011 to 2016.

**Table 1**: Summary of Property crime statistics across seasons and years

|        |        | No of Property crimes | |
|--------|--------|-------|---------|
|        |        | Count | percent |
|        | Summer | 17671 | 21.7 |
|        | Autumn | 20848 | 25.5 |
| Season | Winter | 21248 | 26.0 |
|        | Spring | 21830 | 26.8 |
|        | 2011   | 12922 | 15.8 |
|        | 2012   | 15105 | 18.5 |
|        | 2013   | 637   | 0.8 |
| Year   | 2014   | 17343 | 21.3 |
|        | 2015   | 17970 | 22.0 |
|        | 2016   | 17620 | 21.6 |

There were 81,597 Property crimes in total over the study period: 12922 in 2011, 15105 in 2012, 637 in 2013, 17343 in 2014, 17970 in 2015, and 17620 in 2016. The percentages of Property crimes were

calculated  by dividing Property crimes for each year by the total number of crimes for all years.  Notably, of the 81,597 Property crime cases during the study period: 17,671 were counted in summer, 20,848 in autumn, 21,248 in winter and 21830 in spring.

The number of reported Property crimes ranged from 0 to 258 with an average of 41 cases per year. Findings showed that for the study locations, the minimum number of people per square kilometer was 41 while the maximum is 21 812. In addition, the average number of people per study location was found to be 4611.
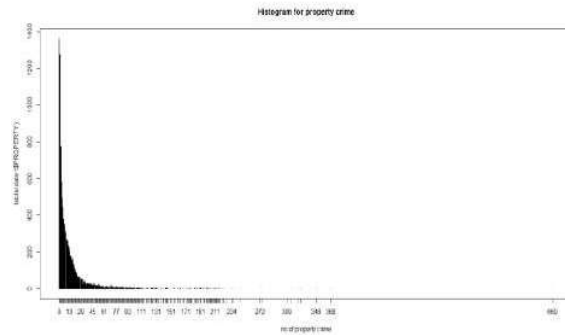


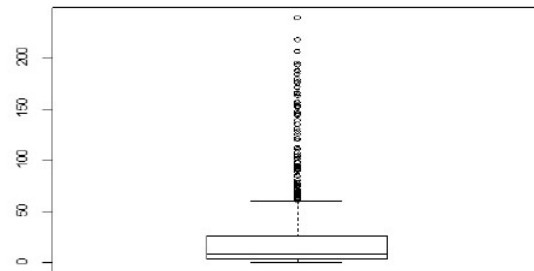**Figure 1**: Histogram for the Number of Property crimes



**Figure 2:** A box plot for the number of property crimes

The histogram (Figure 1) illustrates that the marginal distribution exhibits both substantial variation and a rather large number of zeros. It is clearly evident that the number of property crimes appears to be positively skewed, as indicated by the relative position of the median within the box plot (Figure 2) that contains half the data. However, there are some few outliers as shown in the Figure 2.

There are two distinct processes driving the zeros, one is non-structural zeros (sampling zeros) which occur by chance and can be assumed to be a result of a dichotomous process. The other one is structural zeros (true zeros) which are part of the counting process. Based on this concern, the choice should be based on the model providing the closest fit between observed and predicted values. The choice of the zero-inflated model in this paper is guided by the researcher's belief about the source of the zeros.
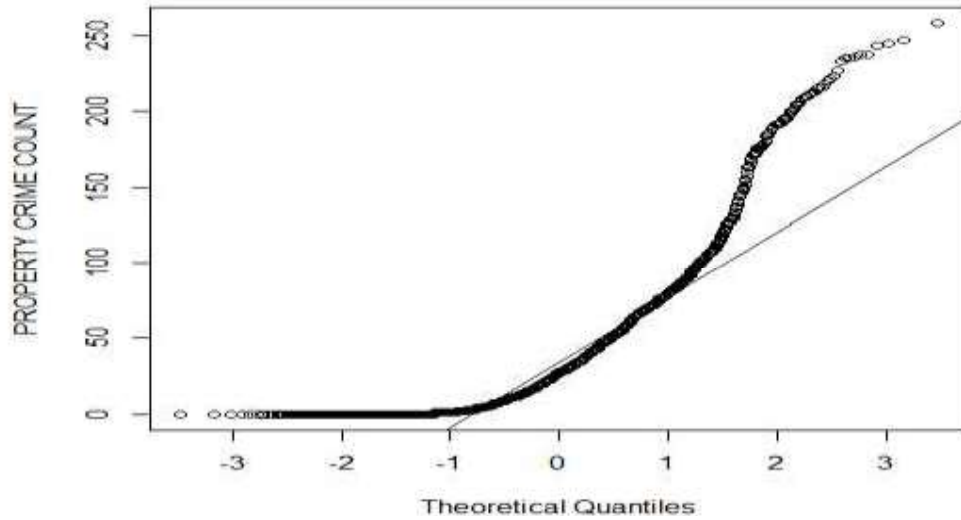
$$Shapiro\ Wilk\ normality\ test\ W\ =\ 0.80677, p-value\ <\ 0.001$$

**Figure 3:** Number of Property crime Q-Qplot.

The Q-Q plot (Figure 3) shows that the Number of Property crimes are positively skewed, as the points fall above the line as x-values increases. This violates a very important assumption for the linear mixed effect model and rather supports generalized linear mixed models. Literature outlined that, linear models are not approriate in some situations where the response is restricted to binary and count. In addition, linear models fail when the variance of the response depends on the mean. Finally, Shapiro's test result indicates that that the distribution is not normal.
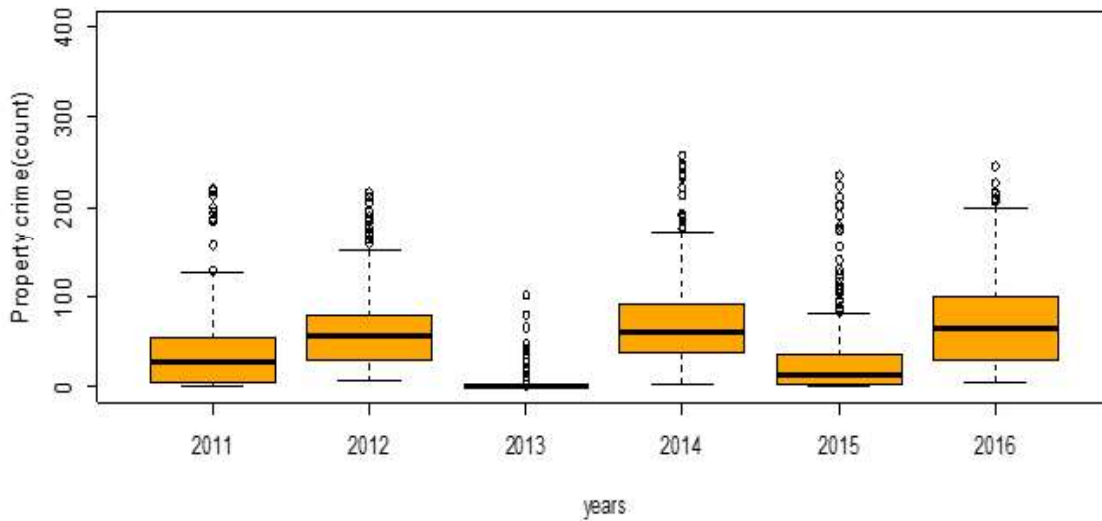


**Figure 4:** Number of Property crime distribution for each year

The boxplot presented in Figure 4, shows the variation in the median Number of Property crimes across the study periods, with the years 2012, 2014 and 2016 showing very similar median of the number of cases. This indicates that the variation across the years needs to be taken into account when fitting the model.
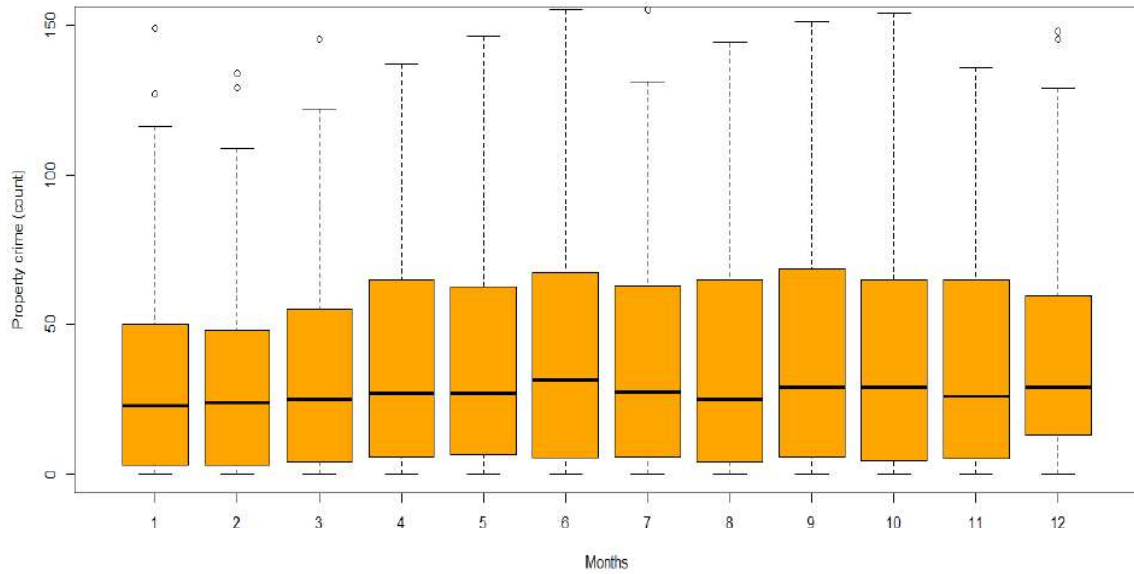
**Figure 5**: The Number of Property crime distribution for each month (January to December).

The boxplot shows that the median for the property crime is slightly different. It also shows that each month presents a different amount of variation in Property crime so that there is an overlap of values between some months. The variation seems to be high as from April to December. There are still noticeable differences and hence the study better accounts for them in the model.
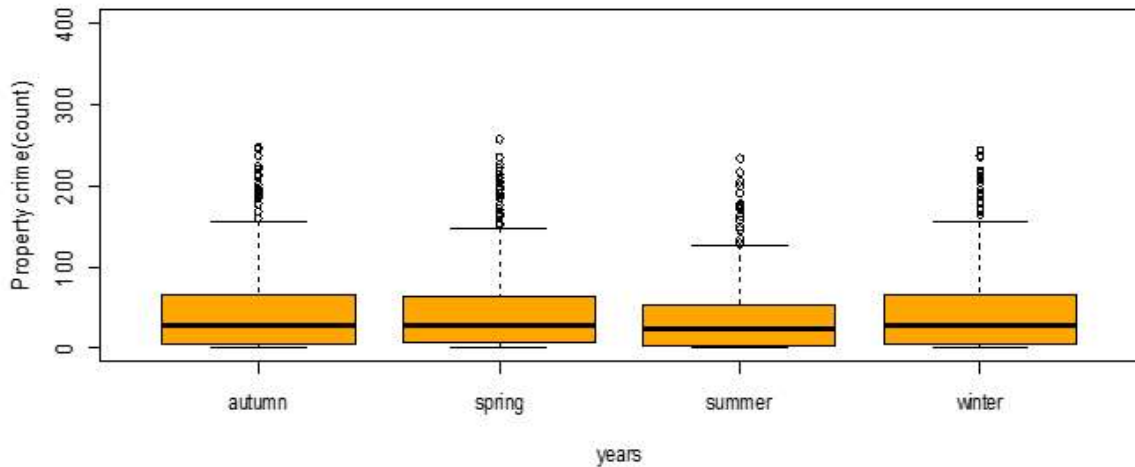


**Figure 6:** Number of Property crimes for each season

The above Figures 4, Figures 5 and Figures 6, demonstrate the possible presence of outliers in the study of property crimes. Hence, suggesting that the study should not heavily rely on the average of property crimes as they may distort the results. It was noted that in all the box plots, the average number of property crimes were above the median since the data was positively skewed.
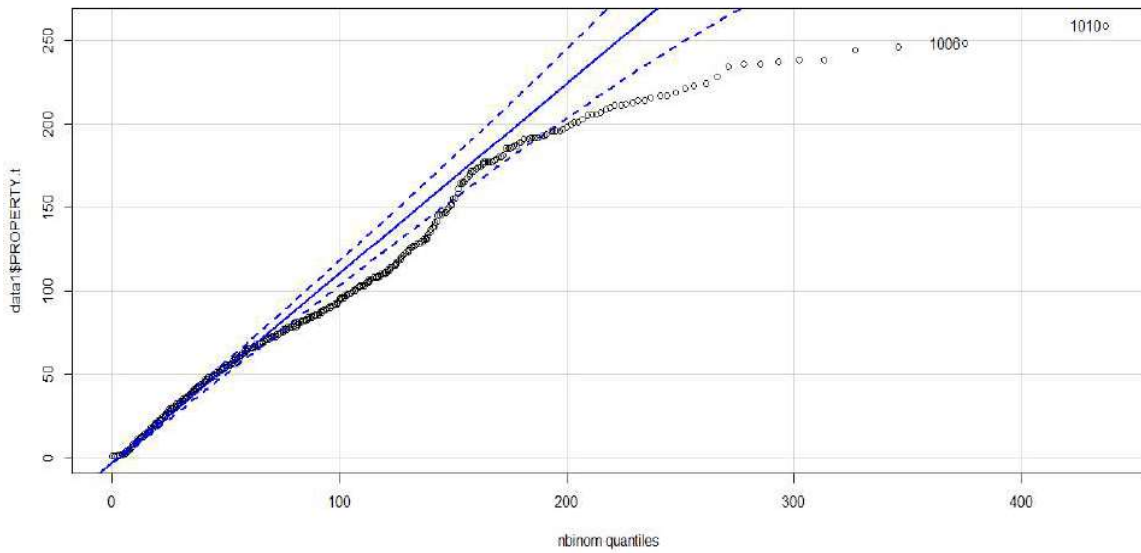
**Figure 7:** Asessing if the Number of Property crime data follows a negative binomial distribution using confidence interval

Considering the plots generated using Q-Q plot in Figure 7, the y-axis represents the observations, and the x-axis represents the quantiles modelled by the distribution. The solid blue line represents a perfect distribution fit, and the dashed blue lines are the confidence intervals of the perfect distribution fit. The aim is to see if the data follows a normal distribution or other distributions. In this case, it is the negative binomial distribution, in which only a few observations fall outside the dashed lines. This suggests that a negative binomial probability distribution best fits the Property crime data.

### 4.1 Model selection

In this study, ZINB-GLMM was reasonable in modelling the Number of Property crimes because of its small AIC values as compared to ZINB, ZIP, NB, and Poisson (Table 2). ZINB-GLMM was the best model for this study based on the benefit that it accounts for within variation through random effects and captures the non-structural zero counts in the dataset.

**Table 2:** Comparing AIC for five Property crime models.

| MODELS | PROPERTY CRIME | |
|---|---|---|
| ZINB-GLMM | 14292.6 | |
| POISSON | 95365 | |
| NB | 18133 | |
| ZIP | 74613 | |
| ZINB | 17972 | |

### 4.2 Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) for the Number Property crime.

For the ZINB-GLMM, the response variable of interest was the Number of Property crimes while the independent variables were *Month*, *Season*, *Location*, *Year* and *Density*. The first four variables (Month, Season, Location and Year ) were the random effects. These effects were chosen to be "random" because the crime commited in one month is independent of the crime commited in the next month. This is also applicable to Location, Season and Year. However, Density was chosen to be the fixed effect since the number of people per square kilometer could be measured during this study period. Moreover, the random effects were nested in this study, because each police officer recorded a certain number of cases, and no two officers recorded the same case.

The full Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) was modelled on the r output format of lme4. The model fitted was:

$$Mixedglmm1 < -glmmTMB(PROPERTY \sim Density2 + (1|Month) + (1|Year) + (1|Location) + (1|Season), zi = \sim Year + disp = \sim Season, data1, family = nbinom2) \quad (4.1)$$

Following the arguments in the function (Mixedglmm1), this model allows the conditional mean to depend on Density and vary randomly by Location $(l)$, Month $(m)$, Year $(y)$ and Season $(s)$. It further allows the number of structural (extra) zeros to depend

on years. Additionally, it allows the dispersion parameter to depend on the season of the year. This model can be represented by the following specific equations (Brooks et al., 2017b):

$$\mu = E(count|a_{lmys}, NSZ) = \exp(\beta_0 + \beta_{Density2} + a_{lmys}) \quad (4.1)$$

$$a_{lmys} \sim N(0, \sigma^2_{a_{lmys}}) \quad (4.2)$$

$$\sigma^2 = var(count|a_{lmys}, NSZ) = \mu(1 + \frac{\mu}{\theta}) \quad (4.3)$$

$$Logit(p) = \beta_0^{z_i} + \beta_{year}^{z_i} \quad (4.4)$$

$$\log(\theta) = \beta_0^{(disp)} + \beta_{season}^{(disp)} \cdot Season \quad (4.5)$$

where $a_{lmys}$ are Location, Month, Year and Season specific random effects, $NSZ$ is the event "non-structure zero", $p = 1 - p_r(NSZ)$ is the zero-inflation probability, and $\beta$'s are regression coefficients with the subscript denoting the covariate/ level (with 0 denoting intercept).

**Table 3:** Summary of the ZINB-GLMM for the Number of Property crimes

| | Parameters | Est. | SE | RR | 95% CI | p-value |
|---|---|---|---|---|---|---|
| **Random effect** | $\sigma_{Location}$ | 1.38 | | | | |
| | $\sigma_{Month}$ | 0.05 | | | | |
| | $\sigma_{Year}$ | 1.05 | | | | |
| | $\sigma_{Season}$ | 0.10 | | | | |
| **Conditional model** | Intercept | 2.91 | 0.30 | 18.31 | 17.72; 18.90 | *** |
| | Density2 | -0.79 | 0.35 | 0.45 | -0.23; 1.13 | * |
| **Zero-Inflation model** | Intercept | -0.50 | 0.36 | 0.60 | -0.11; 1.31 | *** |
| | Ref (2011) | | | | | |
| | 2012 | -17.48 | 2387.26 | 0 | -4679.03; 4679.03 | |
| | 2013 | 5.72 | 0.43 | 304.9 | 304.06; 305.74 | *** |
| | 2014 | -17.53 | 2385.93 | 0 | -4676.42; 4676.42 | |
| | 2015 | -0.77 | 0.50 | 0.46 | -0.52; 1.44 | |
| | 2016 | -17.53 | 2385.93 | 0 | -4676.42; 4676.42 | |
| **Dispersion model** | Intercept | 1.96 | 0.10 | | -0.196; 0.196 | *** |
| | Ref(Summer) | | | | | |
| | Autumn | 0.17 | 0.14 | 1.19 | 0.92; 1.46 | |
| | Winter | -0.11 | 0.14 | 0.9 | 0.63; 1.17 | |
| | Spring | -0.18 | 0.14 | 0.84 | 0.57; 1.11 | |

Estimate (Est.), Standard Error (SE), Relative Risk ratio (OR), Reference category (Ref), Significant level (***=0.001; *=0.05)

The model summary can be broken down into five sections. The first section includes the general overview containing a description of the model specification (family, formula, zero inflation, dispersion, data) together with the information criterion (AIC and BIC).

The second section describes the variability of the random effects. In this model, we only had random effects in the conditional model (equation 4.1). The estimated standard deviations;
$$\sigma_{Location} = 1.38, \quad \sigma_{Month} = 0.05, \quad \sigma_{Year} = 1.05, \quad \sigma_{Season} = 0.10$$
corresponding to $\sqrt{\sigma^2_{a_{lmys}}}$ in equation (4.2). This indicates how much of the variation in the study of property crimes can be attributed to each random term. The variability of the Number of Property crime according to Month and Season was smaller when compared to that of the Location and Year. This was caused by the high dispersion parameter computed under the two variables. The smaller variation in months and seasons, an indication of less time, is

always expected between Property crime cases. To support the view, on the monthly and seasonal perspective, the study has shown that Property crime happens frequently.

The third section describes the relative risk ratio of the conditional model ($\beta_0, \beta_{Density2}$) including a 95% confidence interval and p-values. For the confidence interval, the null value is one since it is estimated on a natural scale. In most cases where a 95% confidence interval does not include the null value the findings are statistically significant. Alternatively, parameters that are statistically significant in the model have a p-value below 0.05 as shown in **Table 3**. Both the intercept ($\beta_0 = 18.31$) and the relative risk ratio $\beta_{Density2} = 0.45$ are statistically significant. Based on this, it means without considering the population density, approximately 18 Property crimes can be anticipated in Windhoek per annum. The expected counts are conditional on every other value being held constant. That is, including the random Location, Month, Year, and Season effects, population density is expected to have a 45 percent increase on property crime. In other

words, considering the population density, the Number Property crimes will increase by 45%.

The fourth section describes the zero-inflation model which is like the conditional model except that this model has a logit link. The estimates in this section correspond to $\beta_0^{z_i}$ and $\beta_{year}^{z_i}$ relative risk ratio from equation (4.4). The Zero-Inflation model estimates the probability of an extra zero. The baseline odds of no property crime reported in Windhoek is 0.6. In addition to that, $\beta_{2013} > 0$ means higher occurrences in year 2013 unaffected by population density. This essentially means that during year 2013, the number of property crimes that were not recorded (but there was an intention) was not due to the number of residents per square kilometre in the area. In contrast, the exploratory study has proven that an area with a high population density experienced high number of property crimes during each season. The confidence intervals were estimated given by:

$$e^{\widehat{\beta_i} \pm Z_{\frac{\alpha}{2}}(S.E(\widehat{\beta_i})} \qquad (4.6)$$

In this case $\widehat{\beta_i}$ represents the model parameters estimated, S.E is the standard error for the corresponding parameter and $Z$ corresponds to the critical value associated with a 95% degree of confidence. The second component of the model [ $Z_{\frac{\alpha}{2}}(S.E(\widehat{\beta_i}))$] is called the margin of error.

Since $\widehat{\beta_i} = \log(mean) = \log(\mu)$, hence $e^{\widehat{\beta_i}} = (mean) = \mu$.

From **Table 3**, the confidence interval for the estimate $\widehat{\beta}_{2013}$ on the conditional model is $304.06 \, ; \, 305.74$. It was concluded that this confidence interval provided the study with plausible values for the parameter. If repeated samples were taken and the 95% confidence interval computed for each sample, 95% of the intervals would contain that population parameter.

## 5. Discussion

During the study period, an average of 68% of the recorded crime was Property crime in Windhoek. This result is not very different from the statistics by NUMBEO (2018) which indicated that Property crimes stood at 74.06%, during the same period.

Specifically, the Number of Property crimes was slightly high during Spring and Winter time. During Spring (October-December) crime increased probably because large numbers of residents normally travel for holidays with their families leaving their houses unattended or with no security. This motivates the offenders to commit Property crimes such as house breaking. Whereas in Winter (July-September), it was assumed that criminals take advantage of the windy

and cold weather to commit property crime as most people prefer to stay indoors due to the cold weather. However, significant to note is that correlation does not mean causation. A possible explanation is not that there are more people on the streets committing crimes during the holidays and Winter times, but also that the possibility of arresting the offenders is lessened by the fact that there could be less cops patrolling in the streets at that time. Also, it is possible that the Property crime rate during these times was higher than what the data shows because of a reporting bias. In general, the number of property crimes has slightly increased in Windhoek between 2011 and 2016.

It was found that there was a direct relationship between the Number Property crime and population densities in Windhoek in line with the study results by (Hanley et al., 2016). The current study results show that most areas with a high population density have a higher crime intensity (with an exemption to Windhoek Central area).

Furthermore, those locations with high population densities were considered as overcrowded areas with people of generally low socio-economic status. Residents in these areas often move around leaving their houses unguarded and this attracts criminals. This is also in agreement with findings by (Cohen & Felson, 2016) that crimes result from the convergence of some elements such as, suitable target, motivated offender, and the absence of capable guardians. Even though Windhoek Central has a small population density more crimes were recorded there as people gather in the area for employment, school, and shopping purposes. Using the same logic, affluent areas attract more criminals for theft and robberies due to the opportunities available to them (Justus & Kassouf, 2013). Besides that, the study results strengthen findings that local crime rate are influenced by income deprivation and housing tenure structures (Livingston, Kearns, & Bannister, 2014).

Although Property crimes were found to be zero inflated, regardless of population density, 18 counts of Property crimes were expected every year. These data revealed that the number of property crimes have been constantly increasing as from 2011 to 2016. The zeros (non- structural zeros) obtained were results of no crime recorded within some months or seasons of the study period due to effective policing or lack of suitable targets and this contributed to the choice of ZINB-GLMM. The study indicated that crime data can be modelled using ZINB-GLMM as an alternative to spatial temporal patterns that were used by other researchers. Perhaps, an important factor to note, is that several other variables were not significant in the model and were not interpreted. However, this is not to say that insignificant findings signal no impact of the indicator on property crimes, since individual level

driving factors cannot be investigated using police data. More complex models that include more variables influencing the Number of Property crime and interaction among these variables at different levels of aggregation would be preferable.

The crime data considered for this study was secondary data received from Windhoek police, focusing only on Windhoek reported crimes. However, due to the sensitive nature of this study, the researcher was given limited information and access to the crime data received from Windhoek police station.

## 6. Conclusion

This research investigated the ecological factors influencing the Number of Property crimes in Windhoek to provide information and insights leading to more effective and improved prevention strategies of these crimes. Premised on this background, the study adopted the ZINB-GLMM approach, which evaluated uncertainty in the random effects contributing to the variation in the Number of Property crimes. The random effects evaluated were based on Month, Year, Season and Location. The results indicated that most of the variation in the study of Property crimes was due to Location while the effect of Month, Year and Season was not as pronounced.

Density was one of the major contributing factors of the high crime rate in Windhoek. Overcrowded areas tend to attract more criminals for theft and robberies due to the opportunities available to them. Okuryangava location was among locations with very high population densities. Even though Hakahana has the highest population density, it was found that the number of reported crimes were quite few.

The study recommends more effective policing in Windhoek during Spring and Winter time, specifically in the areas with high population densities.

Windhoek community members should team up in neighbourhood watch interventions to avoid highly diverse population, with little of the social "glue" that binds communities together, in order to effectively reduce crime. Community members should avoid non-essential mobility and make security arrangements if they have to travel. The Windhoek police could record other important background variables such as the employment status, level of education, age and tribe etc. of the criminal when recording crime to improve data quality in future. It would also be advisable for the Windhoek police to geo-code crime data by location so that future researchers will analyse the spatial aspect of crimes.

## References

Asongu, S., & Kodila-Tedika, O. (2016). Fighting African conflicts and crimes: which governance tools matter? *International Journal of Social Economics, 43*(5), 466-485.

Ayling, J. (2013). What sustains wildlife crime? Rhino horn trading and the resilience of criminal networks. *Journal of International Wildlife Law & Policy, 16*(1), 57-80.

Bolker, B. M. (2018). Getting started with the glmmTMB package.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution, 24*(3), 127-135.

Breetzke, G. D., & Pearson, A. L. (2014). The fear factor: Examining the spatial variability of recorded crime on the fear of crime. *Applied Geography, 46*, 45-52.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017a). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal, 9*(2), 378-400.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017b). Modeling zero-inflated count data with glmmTMB. *bioRxiv*, 132753.

Cohen, L. E., & Felson, M. (2016). Social Change and Crime Rate Trends: A Routine Activity Approach (1979). In *Classics in Environmental Criminology* (pp. 203-232): CRC Press.

Council, O. S. A. (2010). South Africa 2010 Crime & Safety Report. *United States Department of State Bureau of Diplomatic Security*.

de Melo, S. N., Matias, L. F., & Andresen, M. A. (2015). Crime concentrations and similarities in spatial crime patterns in a Brazilian context. *Applied Geography, 62*, 314-324.

Dore, P. (2013). Urbanisation, Inequality and Youth. *State of the Urban Youth, India 2012*, 75.

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics, 5*(3), 236-244.

Ellis, S., & Shaw, M. (2015). Does organized crime exist in Africa? *African Affairs, 114*(457), 505-528.

Flores-Agreda, D., & Cantoni, E. (2019). Bootstrap estimation of uncertainty in prediction for generalized linear mixed models. *Computational Statistics & Data Analysis, 130*, 1-17.

Green, S. B., & Salkind, N. J. (2016). *Using SPSS for Windows and Macintosh, Books a la Carte*: Pearson.

Han, L., Bandyopadhyay, S., & Bhattacharya, S. (2013). Determinants of violent and property crimes in England and Wales: A panel data analysis. *Applied Economics, 45*(34), 4820-4830.

Hanley, Q. S., Lewis, D., & Ribeiro, H. V. (2016). Rural to urban population density scaling of crime and property transactions in English and Welsh parliamentary constituencies. *PloS one, 11*(2), e0149546.

Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ, 2*, e616.

Jones, R. W., & Pridemore, W. A. (2012). The Foreclosure Crisis and Crime: Is Housing-Mortgage Stress Associated with Violent and Property Crime in US Metropolitan Areas? *Social Science Quarterly, 93*(3), 671-691.

Justus, M., & Kassouf, A. L. (2013). Evidence of the effect of wealth observed by criminals on the risk of becoming a victim of property crimes. *Economía, 14*(2), 88-101.

Kemper, J., & ROUX, J. P. (2005). Of squeezers and skippers: factors determining the age at moult of immature African Penguins Spheniscus demersus in Namibia. *Ibis, 147*(2), 346-352.

Khan, N., Ahmed, J., Nawaz, M., & Zaman, K. (2015). The socio-economic determinants of crime in Pakistan: New evidence on an old debate. *Arab Economic and Business Journal, 10*(2), 73-81.

Livingston, M., Kearns, A., & Bannister, J. (2014). Neighbourhood structures and crime: the influence of tenure mix and other structural factors upon local crime rates. *Housing Studies, 29*(1), 1-25.

Magnusson, A., Skaug, H., Nielsen, A., Berg, C., Kristensen, K., Maechler, M., . . . Brooks, M. M. (2017). Package 'glmmTMB'.

Neema, I., & Böhning, D. (2012). Monitoring murder crime in Namibia using Bayesian space-time models. *Journal of Probability and Statistics, 2012*.

NUMBEO. (2018, January 2018). Crime in Windhoek, Namibia. Retrieved from https://www.numbeo.com/crime/in/Windhoek

Palmary, I., Rauch, J., & Simpson, G. (2014). Violent crime in Johannesburg. In *Emerging Johannesburg* (pp. 117-138): Routledge.

Posada, D., & Buckley, T. R. (2004). Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology, 53*(5), 793-808. doi:10.1080/10635150490522304

Schielzeth, H., & Nakagawa, S. (2013). Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution, 4*(1), 14-24.

Shaw, M., & Reitano, T. (2013). The evolution of organised crime in Africa-towards a new response. *Institute for Security Studies Papers, 2013*(244), 28.

Tonry, M. (2014). Why crime rates are falling throughout the Western world. *Crime and justice, 43*(1), 1-63.

Walker, S. (2018). *Crisis-Opportunity, Liability-Asset: Governing Vacant Land Reuse in Cleveland, Ohio.*

Wambua, P. M. (2015). Call the police? Across Africa, citizens point to police and government performance issues on crime.