A SPATIAL-NONPARAMETRIC APPROACH FOR PREDICTION OF CLAIM FREQUENCIES IN MOTOR INSURANCE

Presenter : Kipngetich Gideon

Supervisor 1 : Dr Ananda Kube

Supervisor 2 : Dr Thomas Mageto

Presentation Outline

Background of the study

- Insurance has an important role in providing financial protection and offering a transfer risk in exchange for risk premium.
- Studies on spatial modeling have been rapidly applied in many fields; epidemiology, public health, and the Insurance sector
- Models such as Poisson, Generalized linear models, Credibility models and Bayesian Models are the commonly used models for prediction of claim frequencies.
- Nonparametric models are deemed to minimize the shortcoming of these standard parametric models since fewer assumptions are made for the model.

Background of the study

- Doupe et al. (2019) found that the nonparametric models perform better than generalized linear models (GLMs).
- Fellingham et al. (2015) proposed a Bayesian nonparametric approach for prediction of claims.
- Hong and Martin (2017) also proposed a flexible nonparametric loss model for prediction of the claims
- This study's research idea was to build a spatial nonparametric estimator based on the idea proposed by Wang et al. (2012, 2016) where they introduce a nonparametric regression model with correlated errors.

Statement of the Problem

- Different parametric models have been widely adopted as a standard approach to modeling and predicting claims.
- Approaches rely on restrictive conditions; linearity, normality, and strong distributional assumption on data.
- ANN has been proposed, However, the model has not incorporated the spatial effects in prediction which according to Lee and Durbán (2009) is likely to improve the prediction accuracy.
- However, misspecification and restrictive assumptions on models for prediction in insurance make the results too subjective and less reliable for policy and decision making.
- The study derive a spatial nonparametric regression estimator for prediction of claim frequency

Objectives of the Study

General Objective

- The aim of this study is to derive a spatial nonparametric estimator for estimating claim frequencies in insurance company based on smoothing spline

Specific Objectives

i. To derive a spatial nonparametric estimator for predicting insurance claims frequencies

ii. To investigate theoretical properties of the estimator proposed in (1) above

iii. To apply the estimator proposed in (1) above to a set of claims data of cooperative insurance company

Significance of the study

- The proposed estimator can be used by the insurance companies to predict their claim frequencies.
- The study also contribute to the existing literature on the nonparametric and parametric prediction of claim frequencies by proposing a spatial method for predicting the claims frequencies in insurance industry.

Methodology

- The study proposed a non-parametric regression model to predict the number of claims Y_i, $i = 1, \ldots, n$ observed in region J
- To relax restrictive assumption on the distribution of number of claims and X_i covariates vector for the i^{th} claim.
- Since claims in each region, J have nonlinear relation with the covariates $X_i^{'}s$. The nonparametric form of the model is given by the general form Rice and Wu (2001); Karcher and Wang (2001).

$$y_i = g(x_i) + Z_i^T b + \epsilon_i$$

$Z_i^T b$ and ϵ_i cater for random effects

- Since the form of $Z_i^T b = R_i$ for R_i is unknown. The main work of this study is to estimate the form of R_i that captures the spatial effects.

Cont'd

- We can make assumption about the spatial model as

$$Y_i = g(X_i) + R_i \quad var(R) = \mathbf{\Sigma} \quad i \in \Lambda_n \tag{1}$$

  where $i = (i_1, \ldots, i_N)$ in Λ_n will be referred to as site, R_i cater for the spatial effects
- Spatial data is modelled as finite realization of vector stochastic process indexed by $i \in \Lambda_n$
- $R = (R_1, \ldots, R_n)^T$ is assumed to follow a joint Gaussian distribution where $E(R_i) = 0$, is known $\forall i \in \Lambda_n$
- $\mathbf{\Sigma} = [\rho(R_i, R_j)]$ is the unknown correlation coefficient matrix(need to be estimated). The vector $X_i = (X_{i1}, \ldots, X_{id}) \in \Re^d$, $Y_i \in \Re$ and $g(\cdot)$ is the unknown trend function.
- The aim is to estimate $g(x)$ for some given $x = (x_1, \ldots, x_d) \in \Re^d$ where the response variable Y_i is claim frequency and X_i is six dimensional vector of covariates

Cont'd

- Estimating $g(x)$ at some point $x \in \Re^d$, for X_i in the neighbourhood of x, $\mathbf{g}$ can be approximated using smoothing spline Wang (2019); Tait and Woods (2007).
- To estimate the smoothing spline estimator $\hat{g}(\cdot)$ of $g(\cdot)$, the study consider minimizing the equation

$$\sum_{i=1}^{n} \left(Y_i - g\left(x_i\right)\right)^2 + \lambda \int \left(g''(x)\right)^2 dx$$

  over the function g.
- This criterion trades-off least squares error of $\mathbf{g}$ over $\left(x_i, y_i\right), i = 1, \ldots n$, with a regularization term that grows large when the second derivative of $\mathbf{g}$ is wiggly.
- The coefficients are chosen to minimize the equation (2) which is a simplified form of the above equation

$$\frac{1}{n} \sum_{i=1}^{n} \left\{Y_i - g(X_i; \beta)\right\}^2 + \lambda \beta^T \Omega \beta \tag{2}$$

Cont'd

- which can be represented as

$$\| Y_i - G\beta \|^2 + \lambda \beta^T \Omega \beta$$

$G \in \mathbb{R}^{n \times n}$ is basis matrix defined as

$$G_{ij} = \psi_j \left(x_i \right), \quad i, j = 1, \ldots n$$

$\psi_1, \ldots \psi_n$ are the truncated power basis functions with knots at $x_1, \ldots x_n$ which is evaluated at the data values
$\Omega \in \mathbb{R}^{n \times n}$ is the penalty matrix defined as

$$\Omega_{ij} = \int g_i''(x) \psi_j''(x) dx, \quad i, j = 1, \ldots n$$

- The term affects shrinking the components of estimation $\hat{\beta}$ towards zero. The parameter $\lambda \geq 0$ is the smoothing parameter.

Cont'd

- Given the optimal coefficients $\hat{\beta}$ minimizing (2) through penalized least squares, the smoothing spline estimator at x is therefore defined as

$$\hat{g}(x_i) = \sum_{j=1}^{n} \hat{\beta}_j \psi_j(x) \tag{3}$$

- Each computed coefficient $\hat{\beta}_j$ corresponds to a particular basis function ψ_j.
- The term $\beta^T \Omega \beta$ in (2) imparts more shrinkage on the coefficients $\hat{\beta}_j$ that correspond to wigglier functions $\psi_j(x)$. Hence, as we increase λ, we are shrinking away the wiggler basis functions
- Similar to least squares regression, the coefficients $\hat{\beta}$ minimizing (2) is

$$\hat{\beta} = \left(G^T G + \lambda \Omega\right)^{-1} G^T Y = (X^T X + n\lambda D)^{-1} X^T Y$$

Cont'd

- where X is a design matrix with entries x_i for $i = 1, \ldots, n$, Y is a vector of the response variables, D is a diagonal matrix with $p + 1$ zeros on the diagonal followed by N ones and $n\lambda D$ is a penalty term
- Smoothing splines can be seen as a linear smoother, where $k(x) = (\psi_1(x_1), \ldots \psi_n(x_n))$. Therefore, equation (4) can be represented as

$$\hat{g}(x) = k(x)^T \hat{\beta} = k(x)^T (X^T X + n\lambda D)^{-1} X^T Y \tag{4}$$

which is linear combination of the points $y_i, i = 1, \ldots n$

- λ is estimated using Generalized Cross Validation (GCV) method given by

$$GCV(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{(Y(z_i) - \hat{Y}_\lambda^{-i}(z_i))}{1 - (p + tr(S_\lambda))/n} \right)^2 \tag{5}$$

where $Y(z_i)$ is the observation in point z_i, $Y_\lambda^{-i}(z_i)$ is the predicted value from a fitted smoothing spline model from the data and S_λ is the degree of the smoother

Cont'd

- As proposed by Wang et al. (2012, 2016), R^2 is used to assess the performance of predictor function, given by

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[g(x_i) - \hat{g}(x_i)]^2}{\sum_{i=1}^{n}[g(x_i) - \bar{g}]^2} \tag{6}$$

  where $\bar{g}$ is the sample mean of $g(x_i), \quad i = 1, \ldots, n$.
- After estimating the function $g(\cdot)$, then from (1) R_i is estimated as $\hat{R}_i = Y_i - \hat{g}(X_i)$.
- Since $\mathbf{\Sigma}$ in model equation (1) is unknown, we assume that $R_i, i = 1, 2, \ldots, n$ is $2^{nd}-$ order stationary and isotopic process (does not depend on direction)
- Let $C(h)$ and $2\gamma(h)$ be covariogram and variogram of the process where h represents the distance between 2 points at which the process is obtained Laslett (1994); Wang et al. (2017).

Cont'd

- The two quantities are related by

$$C(h) = C(0) - \gamma(h) \tag{7}$$

where $C(0) = \sigma^2 = var(Y(\boldsymbol{z}))$, $Y(\boldsymbol{z})$ is the value of the process at spatial location $\boldsymbol{z}$ within region $\boldsymbol{C}$.

$$\lim_{h \longrightarrow \infty} C(h) = 0$$

implies

$$\lim_{h \longrightarrow \infty} \gamma(h) = Var(Y(\boldsymbol{z})) = C(0)$$

for validity of variogram the condition that

$$\lim_{h \longrightarrow \infty} \frac{2\gamma(h)}{h^2} = 0$$

must be met Cressie (2015).
$\sum = [\rho(R_i, R_j)] = [C(\| z_i - z_j \|)/\sigma^2]$, while z_i and z_j are the spatial locations associated with the error values R_i and R_j

Cont'd

- Thus to estimate $\boldsymbol{\Sigma}$ it is sufficient to estimate $\gamma(h)$ Huang et al. (2011); Cressie (2015).

$$2\hat{\gamma}(h) = \sum_{S(h)} [z_i - z_j]^2 / N(h) \tag{8}$$

  $S(h) = \{(z_i, z_j) :\mid z_i - z_j \mid = h\}$, $h \in \Re^d$, $N(h)$ is a number of distinct pairs in $S(h)$
  since $r(z_i)$ the error at location z_i is unobserved, the quantity is to be estimated as well.

- Since we have to estimate the variogram $\hat{\gamma}(h)$ in equation (9) in nonparametric approach Fernández-Casal et al. (2018); Qadir and Sun (2020), then $\gamma(h)$ can be estimated as

$$\gamma(h) = \int_0^\infty (1 - \omega_d(ht))dM(t) \tag{9}$$

Cont'd

- $M(t)$ is nonnegative bounded nondecreasing function for nodes(or location of the jumps) $t \geq 0$ and ω_d is a basis for functions in $\mathbb{R}^d$ (d is the dimension of the spatial domain D) given by

$$\omega_d(ht) = (2/ht)^{(d-2)/2}\Gamma(d/2)J_{(d-2)/2}(ht)$$

  $\Gamma(d/2)$ is the gamma function, and $J_{(.)}$ is the Bessel function of the first kind.

- The characteristics of the estimator (10) is estimated using Integrated square error Yu et al. (2007), given by

$$ISE(\gamma) = \int_{h_1}^{h_k} \{\hat{\gamma}(h) - \gamma(h)\}^2 dh \qquad (10)$$

Cont'd

- Where h_1 and h_k are the smallest and largest distances for which variogram estimates are available Huang et al. (2011).
- Model (1) can therefore be represented as

$$Y\left(\boldsymbol{z}_i\right) = g\left(\boldsymbol{X_i}\left(\boldsymbol{z}_i\right)\right) + \boldsymbol{R}\left(\boldsymbol{z}_i\right), \quad i = 1, \ldots, n \tag{11}$$

where $Y(z_i) : i = 1, \ldots, n$ is the observations in region z_i associated with independent variables $X_i(z_i)$ in region z_i, $R(z_i)$ is the unobserved error in region z_i

- To evaluate performance of the proposed method we used R^2 to assess prediction accuracy of the method

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left[Y\left(z_i\right) - \hat{Y}\left(z_i\right)\right]^2}{\sum_{i=1}^{n}\left[Y\left(z_i\right) - \bar{Y}\right]^2} \tag{12}$$

Theoretical properties of Estimator

- To obtain asymptotic results, we impose the following assumptions on model (1)
  1. The function $g(\cdot)$ is twice differentiable and Its matrix of second derivatives at x denoted by $g^{''}(\cdot)$ is continuous at all $x \in \Re^d$
  2. The process $R_i, i = 1, 2, ..., n$ is 2^{nd}order stationary and isotopic, further $\exists$ a $\epsilon > 0$ such that $E(\mid Y_i \mid^{2+\epsilon}) < \infty$ for $i = 1, 2, \ldots, n$

Theorem 1: Asymptotic Normality

- Suppose that $\{\epsilon_i\}_{i=1}^{n}$ are iid with mean 0 and variance $\sigma^2 I_n$ then, for any $x_i \in \mathbb{R}$ and $k_o \geq C_n^{1/2m+1}$ for some constant $C > 0$ then

$$\frac{\hat{g}(\cdot) - (g(\cdot) + b(x_i))}{\sqrt{var(\hat{g}(\cdot))}} \longrightarrow_d N(0,1) \quad as \quad n \longrightarrow \infty \tag{13}$$

where $b(x_i)$ is the asymptotic bias Shen et al. (1998), given by

$$b(x_i) = E(\hat{g}(x_i) - g(x_i) = b(x_i)$$

Cont'd

- **Proof**:
- For $m > 1$ denote $S(m, \mathbf{t})$ to be the set of spline functions with fixed knots $\mathbf{t} = \{0 = t_0 < t_1 < \cdots < t_{k+1} = 1\}$ of step functions with jumps at the knots and for $m \geq 2$

$$S(m, \mathbf{t}) = s \in C^{m-2}[0, 1] : s(x) \quad is \quad a \quad polyomial \quad of \quad degree(m-1)$$

- Expressing functions in $S(m, \mathbf{t})$ in terms of B-splines for any fixed m and $\mathbf{t}$, let

$$R_{i,m}(x) = (t_i - t_{i-m})[t_{i-m}, \ldots, t_i](t - x)_+^{m-1}, i = 1, \ldots, k = k_0 + M$$

- where $[t_{i-m}, \ldots, t_i]g$ denote the $m^{th} - order$ divide difference of the function g and $t_i = t_{min(max(i,0),k_0+1)}$ for any $i = 1 - M, \ldots, k$ we assume that

$$max_{1 \leq i \leq k_0} |h_{i+1} - h_i| = o(k_0^{-1}) \tag{14}$$

and $h/min_{1 \leq i \leq k_0} h_i \leq M$, where $h_i = t_i - t_{i-1}, h = max_{1 \leq i \leq k_0} h_i$ and $M > 0$ (predetermined constant)

Cont'd

- This assumption ensure that $M^{-1} < k_0 h < M$, which is necessary for numerical assumptions
- Let $D_n(x)$ be an empirical distribution function of $(x_i^n)_{i=1}$ with a positive continuous density $d(x)$ this implies
  $G(d) = \int_0^1 R(x)R'(x)d(x)dx$
- Then
$$\frac{E(\hat{g}(x))}{\sqrt{var(\hat{g}(x))}} - \frac{g(x)+b(x)}{\sqrt{var(\hat{g}(x))}} = \frac{o(k_0^{-m})}{\sqrt{k_0/n}} = o(n^{1/2}k_0^{-(m+1/2)}) = o(1)$$
- Thus equation (14) follows if
$$\frac{\hat{g}(x) - E(\hat{g}(x))}{\sqrt{(\hat{g}(x))}} \longrightarrow_d N(0,1)$$
  we have
$$\hat{g}(x) - E(\hat{g}(x)) = R'(x)G_{k,n}^{-1}x_\epsilon = \sum_{i=1}^{n} a_i\epsilon_i \tag{15}$$

Cont'd

- Where $a_i = R^{'}(x)G_{k,n}^{-1}R(x_i)/n$, the required Lindeberg-Feller conditions, it suffices to verify that

$$max_{1\leq i\leq n}(a_i^2) = o\left(\sum_{i=1}^{n} a_i^2\right) = o(var(\hat{g}(x))) \tag{16}$$

- Finally, equation (15) follows from the assumption that $k_0/n \longrightarrow 0, \quad hn \longrightarrow \infty$ hence the prove.

Theorem 2: Consistency

- From theorem 1, we can establish the consistency of $\hat{g}(\cdot)$ where for $\forall\epsilon > 0$

$$\lim_{n\to\infty} P\left\{\mid \hat{g}(x) - g(x) \mid > \epsilon\right\} = 0 \tag{17}$$

proof:
The mean squared error of $\hat{g}(x)$ is given by
$MSE(\hat{g}(x)) = var(\hat{g}(x)) + [bias(\hat{g}(x))]^2$

Cont'd

- From equation (14) and $\sup_{x\in[0,1]} \mid D_n(x) - D(x) \mid = o(k_o^{-1})$

$$var(\hat{g}(x)) = \frac{\sigma^2}{n} R^{'}(x) G^{-1}(d) R(x) + o((nh)^{-1}),$$

$$E(\hat{g}(x)) - g(x) = b(x) + o(h^m) \quad b(x) = -\frac{f^{(m)}(x) h_i^m}{m!} B_m \left(\frac{x - t_i}{h_i} \right)$$

where $B_o(x) = 1,\ B_i(x) = \int_0^x iB_{i-1}(z)dz + b_i$ and $b_i = i \int_0^1 \int_0^x B_{i-1}(z)dzdx$ is the i^{th} Bernoulli number Barrow et al. (1978)

- From equation (18)

$$\lim_{n\to\infty} P\left\{\mid \hat{g}(x) - g(x) \mid > \epsilon\right\} \leq \lim_{n\to\infty} \frac{MSE(\hat{g}(x))}{\epsilon^2}$$

Cont'd

$$\leq \lim_{n\to\infty} \frac{\frac{\sigma^2}{n} R'(x) G^{-1}(d) R(x) + o((nh)^{-1})}{\epsilon^2} + \frac{\left[-\frac{f^{(m)}(x) h_i^m}{m!} B_m \left(\frac{x - t_i}{h_i} \right) + o(h^m) \right]^2}{\epsilon^2} \tag{18}$$

as $n \to \infty$ the numerator terms in right hand side of equation (19) tends to zero therefore

$$\leq \lim_{n\to\infty} \frac{\frac{\sigma^2}{n} R'(x) G^{-1}(d) R(x) + o((nh)^{-1})}{\epsilon^2} + \frac{\left[-\frac{f^{(m)}(x) h_i^m}{m!} B_m \left(\frac{x - t_i}{h_i} \right) + o(h^m) \right]^2}{\epsilon^2} = 0$$

Hence the prove of equation (18)

Simulation Study

- We simulated spatial data with a length of n=100 observations.
- 65 spatial sampling locations were selected randomly and denoted by $\mathbf{z}_1, \ldots, \mathbf{z}_n$.
- The responses $Y(\mathbf{z}_i)$ for $i = 1, \ldots, n$ are the observations and were simulated from the spatial nonparametric model (12) with $p = 2$

$$Y(\mathbf{z}) = g(x_i(\mathbf{z})) + R(\mathbf{z})$$

  $R(z)$ is the term for spatial effects z_i and z_j in 2-dimensional space with mean 0 and covariance given by (9).
- The covariates, $x_i(\mathbf{z}_i)$ for $i = 1, \ldots, n$, were generated as iid $N(0, 1)$ and are independent of each other

Cont'd

- Within each simulation, the spatial random effects $R(\mathbf{z}_i)$ were generated from a Gaussian process with mean zero and the covariance function (9), for $i = 1, \ldots, n$.
- The ISE(γ) defined as $ISE(\gamma) = \int_{h_1}^{h_k} \{\hat{\gamma}(h) - \gamma(h)\}^2 dh$ was approximated numerically from simulated data for the proposed estimator (9) and NW kernel.
- From Table 1, the proposed estimator (9) offers a better performance compared to NW kernel estimator

Table 1: Mean values of the standardized ISE from the estimators(Sample size n = 1000)

	h=1	h=2	h=3	h=4
NW kernel	0.53883	0.53407	0.5177	0.40779
Proposed estimator	0.30731	0.28762	0.26737	0.23037

Cont'd

- We compare the proposed method under which $\gamma(h) = \int_0^\infty (1 - \omega_d(ht))dM(t)$ with the method under which the spatial component (R) is based on kernel estimation
- We calculated the MSE and the R^2 of the estimators from 100 simulations and present the results in the Table 2

Table 2: MSE and R^2 for the model over 100 simulations

	n=10		n=100		n=400		n=1000	
	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2
$N(kernel)$	0.0308	0.6751	0.0285	0.7585	0.0210	0.9691	0.0176	0.9694
Proposed Method	0.0221	0.7003	0.0118	0.8217	0.0105	0.9962	0.0102	0.99963

Cont'd

- MSE for the proposed method ranges from 0.0221 to 0.0102 while the MSE of the kernel based estimator ranges between 0.308 to 0.0176.
- R^2 for the proposed method ranges between 0.7003 to 0.99963 compared to R^2 of kernel based estimator which rages from 0.6751 to 0.9694.

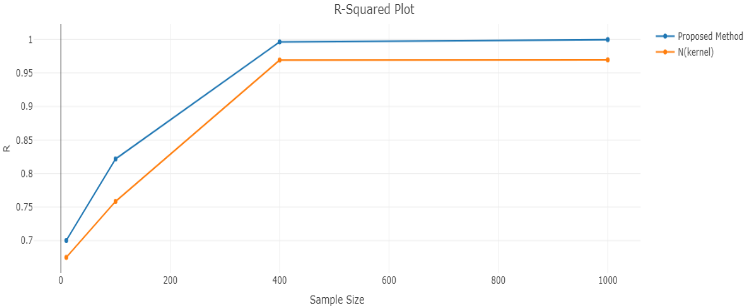

Figure 1: R-squared Plot

Cont'd

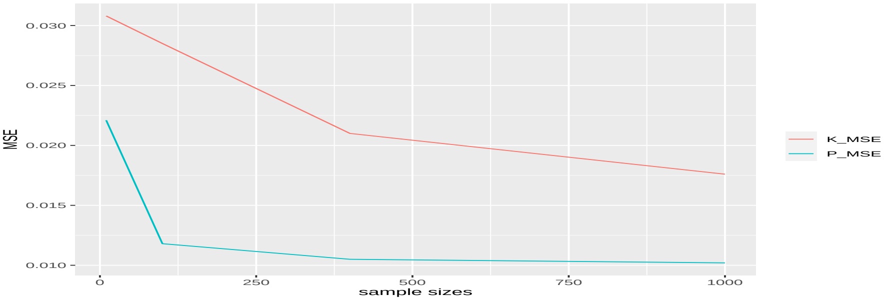

Figure 2: Mean Squared Plot of both K(kernel) and proposed estimator

- The method was applied to the simulated data to check its performance in prediction.
- Table 3 describe the distribution of the predicted values

Cont'd

Table 3: Summary of predicted claim frequencies from simulation

No. of Claims	Freq of Observations	% of Observations
1	998	99.8
2	2	0.2

- The prediction interval (in red dotted lines) showed that a larger number of predicted values lies between $1 - 2$

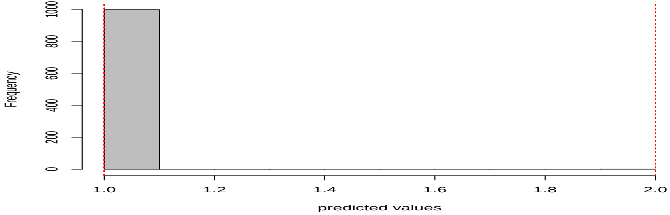

Figure 3: Histogram for the predicted values

Analysis for Claims Data

- The study considered claims data from CIC insurance observed in different parts of 7 counties of Kenya to exhibit the performance of the proposed method.
- The main interest of this study was predicting claims frequencies, the study considers a set of 6500 observations.
- Let Y_i denote the claim frequency, and $\boldsymbol{X_i} = (X_1, \ldots, X_6)^\top$ be a vector which consists of the following explanatory variables: gender, claim amount, age of the policyholder, gender, vehicle age, model of the vehicle and age category of the policyholder.
- Using the estimated model (12) we predicted claim frequencies.

Cont'd

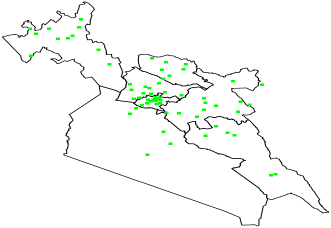

Figure 4: Hotspots locations in Nakuru, Nairobi, Kajiado, Muranga, Kiambu, Machakos, Makueni counties

Test for spatial correlation

- The presence of spatial correlation was determined, figure 4 showed that claims have a significant level of spatial correlation out to distances of about 250 KM.

Cont'd

- The correlation collapses, indicating that as the distances grows bigger the spatial correlation on the claims becomes Zero.

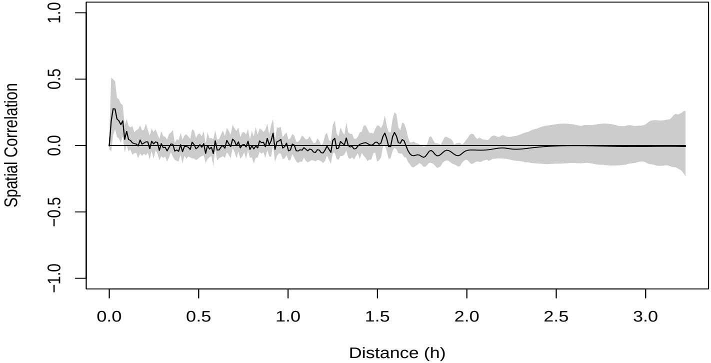

Figure 5: Spatial Correlation

Cont'd

- Using the proposed method, claim frequencies were predicted and the results presented in Table 4

Table 4: Summary of predicted claim frequencies

No. of Claims	Freq of Observations	% of Observations
1	3145	89.52
2	302	8.60
3	50	1.42
4	16	0.50

Figure 6 shows the histogram for the predicted claims, the future values will lie between 1 and 4.

Cont'd

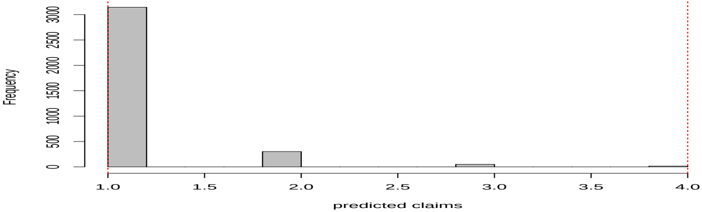

Figure 6: Predicted number of claims

From the prediction results, R^2 values using equation (13) were evaluated to access the performance of two method

Table 5: R^2 for the estimators.

Method	$N(Kernel)$	Proposed Method
R^2	0.543	0.566

Cont'd

- From Table 5, R^2 for $N(Kernel)$ was 0.543, and that from the proposed method is 0.566.
- This showed that the proposed method for prediction has a higher prediction accuracy than the kernel based estimator.
- The study concluded that the proposed method is more efficient than $N(Kernel)$ model

Conclusion and Recommendation

Conclusion

- The simulation study showed that the proposed method outperform the kernel based estimator
- Case study findings also showed that the proposed method outperform the kernel based estimator on predicting the claim frequencies.

Summary

- The proposed method compared to kernel based estimator provides a more efficient prediction method of the motor insurance claim data and ultimately leads to more accurate predictions.
- We expect this study will help to utilize the spatial nonparametric estimators and improve them accordingly to suit the needs for the emerging issues in risk management in insurance.

Cont'd

Recommendations

- From the findings, the study recommends the use of this proposed method in prediction of the claim frequencies in insurance.
- This study make assumption that the errors are correlated for this reason future studies could consider a case of uncorrelated errors.

References

Barrow, D. L., Chui, C. K., Smith, P. W., and Ward, J. D. (1978). Unicity of best mean approximation by second order splines with variable knots. *Mathematics of Computation*, 32(144):1131–1143.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Doupe, P., Faghmous, J., and Basu, S. (2019). Machine learning for health services researchers. *Value in Health*, 22(7):808–815.

Fellingham, G. W., Kottas, A., and Hartman, B. M. (2015). Bayesian nonparametric predictive modeling of group health claims. *Insurance: Mathematics and Economics*, 60:1–10.

Fernández-Casal, R., Castillo-Páez, S., and Francisco-Fernández, M. (2018). Nonparametric geostatistical risk mapping. *Stochastic Environmental Research and Risk Assessment*, 32(3):675–684.

Hong, L. and Martin, R. (2017). A flexible bayesian nonparametric model for predicting future insurance claims. *North American Actuarial Journal*, 21(2):228–241.

Huang, C., Hsing, T., and Cressie, N. (2011). Nonparametric estimation of the variogram and its spectrum. *Biometrika*, 98(4):775–789.

Karcher, P. and Wang, Y. (2001). Generalized nonparametric mixed effects models. *Journal of Computational and Graphical Statistics*, 10(4):641–655.

Laslett, G. M. (1994). Kriging and splines: an empirical comparison of their predictive performance in some applications. *Journal of the American Statistical Association*, 89(426):391–400.

Lee, D.-J. and Durbán, M. (2009). Smooth-car mixed models for spatial count data. *Computational Statistics & Data Analysis*, 53(8):2968–2979.

Qadir, G. A. and Sun, Y. (2020). Semiparametric estimation of cross-covariance functions for multivariate random fields. *Biometrics*.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57(1):253–259.

Shen, X., Wolfe, D. A., and Zhou, S. (1998). Local asymptotics for regression splines and confidence regions. *The annals of statistics*, 26(5):1760–1782.

Tait, A. and Woods, R. (2007). Spatial interpolation of daily potential evapotranspiration for new zealand using a spline model. *Journal of Hydrometeorology*, 8(3):430–438.

Wang, H., Lin, J., and Wang, J. (2016). Nonparametric spatial regression with spatial autoregressive error structure. *Statistics*, 50(1):60–75.

Wang, H., Wang, J., and Huang, B. (2012). Prediction for spatio-temporal models with autoregression in errors. *Journal of Nonparametric Statistics*, 24(1):217–244.

Wang, H., Wu, Y., and Chan, E. (2017). Efficient estimation of nonparametric spatial models with general correlation structures. *Australian & New Zealand Journal of Statistics*, 59(2):215–233.

Wang, Y. (2019). *Smoothing splines: methods and applications*. Chapman and Hall/CRC.

Yu, K., Mateu, J., and Porcu, E. (2007). A kernel-based method for nonparametric estimation of variograms. *Statistica Neerlandica*, 61(2):173–197.

THANK YOU FOR YOUR ATTENTION