



Proceedings of the 64th Annual Conference of the South African Statistical Association for 2023

29 November –
1 December 2023
Durban



Proceedings of the 64th Annual Conference of the South African Statistical Association for 2023 (SASA 2023)

ISBN 978-0-7961-3746-3

Editor

Charl Pretorius North-West University

Assistant Editors

Sugnet Lubbe Stellenbosch University
Andréhette Verster University of the Free State

Managing Editor

Charl Pretorius North-West University

Review Process

Eight (8) manuscripts were submitted for possible inclusion in the Proceedings of the 64th Annual Conference of the South African Statistical Association. All submitted papers were assessed by the editorial team for suitability, after which all papers were sent to be reviewed by two independent reviewers each. Papers were reviewed according to the following criteria: relevance to conference themes, relevance to audience, standard of writing, originality and critical analysis. After consideration and incorporation of reviewer comments, four manuscripts were judged to be suitable for inclusion in the proceedings of the conference.

Reviewers

The editorial team would like to thank the following reviewers:

Renette Blignaut	University of the Western Cape
Jan Blomerus	University of the Free State
Warren Bretteny	Nelson Mandela University
Humphrey Brydon	University of the Western Cape
Allan Clark	University of Cape Town
Legesse Debusho	University of South Africa
Tertius de Wet	University of Stellenbosch
Victoria Goodall	VLG Statistical Services
Gerrit Grobler	North-West University
Johané Nienkemper-Swanepoel	Stellenbosch University
Ibidun Obagbuwa	Sol Plaatje University
Etienne Pienaar	University of Cape Town
Gary Sharp	Nelson Mandela University
Neill Smit	North-West University
Vaughan van Appel	University of Johannesburg
Sean van der Merwe	University of the Free State
Stephan van der Westhuizen	Stellenbosch University
Tanja Verster	North-West University

Contact Information

Queries can be sent by email to the Managing Editor (managing.editor@sastat.org).

Table of Contents

Directional Gaussian spatial processes for South African wind data <i>J. S. Blom, P. Nagar and A. Bekker</i>	1
Information transmission between Bitcoin and other asset classes on the Johannesburg Stock Exchange <i>K. Els, C. Mills, W. Turkington and C.-S. Huang</i>	13
A comparative study of ridge-based adaptive weights in penalised quantile regression on variable selection and regularisation <i>I. Mudhombo and E. Ranganai</i>	27
Bandwidth selection in a generic similarity test for spatial data when applied to unmarked spatial point patterns <i>J. Nel, R. Stander and I. N. Fabris-Rotelli</i>	43

Directional Gaussian spatial processes for South African wind data

*Jacobus S. Blom*¹, *Priyanka Nagar*² and *Andriëtte Bekker*^{1,3}

¹Department of Statistics, University of Pretoria, Pretoria, South Africa

²Department of Statistics and Actuarial Science, Stellenbosch University, Stellenbosch, South Africa

³Centre of Excellence in Mathematical and Statistical Sciences, Johannesburg, South Africa

Accurate wind pattern modelling is crucial for various applications, including renewable energy, agriculture, and climate adaptation. In this paper, we introduce the wrapped Gaussian spatial process (WGSP), as well as the projected Gaussian spatial process (PGSP) custom-tailored for South Africa's intricate wind behaviour. Unlike conventional models struggling with the circular nature of wind direction, the WGSP and PGSP adeptly incorporate circular statistics to address this challenge. Leveraging historical data sourced from meteorological stations throughout South Africa, the WGSP and PGSP significantly increase predictive accuracy while capturing the nuanced spatial dependencies inherent to wind patterns. The superiority of the PGSP model in capturing the structural characteristics of the South African wind data is evident. As opposed to the PGSP, the WGSP model is computationally less demanding, allows for the use of less informative priors, and its parameters are more easily interpretable. The implications of this study are far-reaching, offering potential benefits ranging from the optimisation of renewable energy systems to the informed decision-making in agriculture and climate adaptation strategies. The WGSP and PGSP emerge as robust and invaluable tools, facilitating precise modelling of wind patterns within the dynamic context of South Africa.

Keywords: Directional statistics, MCMC, Projected Gaussian spatial process, Sustainable Development Goal 7, Wrapped Gaussian spatial process.

1. Introduction

The objectives, tactics, long-term aspirations, and growth trajectory pertaining to renewable energy under the framework of Sustainable Development Goal 7 (SDG-7) in the United Nations' 2030 Sustainable Development Goals¹ (SDGs) are designed to facilitate universal access to power, clean cooking fuels, and advanced technologies. A concise overview of the latest findings and methodologies pertaining to the conversion of energy derived from renewable sources into usable forms is presented by Trinh and Chung (2023). Over the past decade, there has been a notable growth in the proportion of the worldwide population that has obtained access to electricity, marking a significant milestone. However, it is worth noting that the number of individuals lacking access to electricity

Corresponding author: Priyanka Nagar (pnagar@sun.ac.za)

MSC2020 subject classifications: 62H11, 62P12, 62R30

¹<https://sdgs.un.org/goals>[accessed 31 October 2023]

in Sub-Saharan Africa has experienced a concerning rise during the same period². South Africa must take measures toward the implementation of renewable energy initiatives in a global context where the popularity of fossil fuels is waning and climate action is viewed as an absolute necessity. Wind power could provide a remedy to South Africa's persistent energy shortages. Nevertheless, harnessing wind energy is a complex endeavour that requires a nuanced understanding of a variety of factors. The study of wind energy holds significant relevance in promoting the four key aspects of energy access, energy efficiency, renewable energy, and international collaboration, hence facilitating the advancement of Sustainable Development Goals. Therefore, modelling wind patterns is crucial in modern society for multiple reasons, including renewable energy, weather forecasting, air quality, and aviation.

Numerical models for weather forecasts require statistical post-processing. Linear variables such as wind speed can be post-processed in different ways as shown in Jona-Lasinio et al. (2007), Kalnay (2002) and Wilks (2006), whereas a circular (or angular) variable like wind direction cannot be post-processed using standard methods (Engel and Ebert, 2007; Bao et al., 2010). Bias correction and ensemble calibration techniques for determining the direction of wind are discussed in Bao et al. (2010). For the bias correction, Bao et al. (2010) considered a circular-circular regression model as proposed in Kato et al. (2008) and for the ensemble calibration a Bayesian model averaging with the von Mises distribution was considered. However, this study did not consider the spatial configuration in the data. The challenge is incorporating structured dependence into directional data. Directional statistics have been developed for many years now starting as early as 1961, where the authors studied complex circular distributions underlying the theoretical framework (Watson, 1961; Stephens, 1963; Kent, 1978). Various approaches to dealing with circular data, distribution theory and inference can be found in Ley and Verdebout (2017), Jupp and Mardia (2009) and Mardia (1972). Previous studies conducted by Rad et al. (2022) and Arashi et al. (2020) explore the feasibility of predicting wind direction in South Africa. Nevertheless, the inclusion of the spatial component in these studies was also lacking.

In the past, spatial models were employed to model wind patterns, but they had challenges with accounting for wind's nonlinear and complicated behaviour. Due to the spatial dependence structure that arises in wind data, a straightforward linear model cannot be used to model wind patterns, as discussed in Jona-Lasinio et al. (2012). Coles (1998) proposed a wrapped Gaussian model for modelling wind directions. The approach assumed an unspecified covariance matrix and independent angular information, working in low dimensions. However, an extension to a spatial framework was briefly discussed. This extension was later introduced by Casson and Coles (1998) where the circular variables were considered to be conditionally independent von Mises distributed. More recently, Jona-Lasinio et al. (2012) introduced a model to analyse wave direction data using a wrapped Gaussian spatial process (WGSP). The WGSP takes into account the spatial structure of directional variables with a potential for high-dimensional multivariate observations which are driven by a spatial process. The methodology allows for the implementation of spatial prediction of the mean direction and concentration while also capturing the dependence structure.

In this paper, we consider the WGSP and projected Gaussian spatial process (PGSP) for modelling wind patterns in South Africa. These models account for the highly complex dependence structure

²<https://www.iea.org/reports/sdg7-data-and-projections/access-to-electricity>[Accessed 31 October 2023]

that arises in wind data as well as the periodic nature of directional data as developed by Jona-Lasinio et al. (2012) (See also Ley and Verdebout, 2018). There are significant distinctions between the two approaches. The wrapping approach constructs a circular distribution that is similar (generally) to its real line counterpart. In other words, if the real line distribution is symmetric and unimodal then the wrapped distribution will have the same characteristics (Jammalamadaka and SenGupta, 2001). The projected Gaussian model, however, may result in differing characteristics from the real line counterpart. For example, the projected Gaussian model can be asymmetric and bimodal. The main justification for proposing these two techniques resides in that it is simple to introduce spatial dependence. The wrapping produces results that are relatively simple to interpret in terms of phenomenon behaviour, whereas the projection is extremely useful when interpretation is less critical and a highly flexible model is required, as stated in Mastrantonio et al. (2016).

The remainder of the paper is structured as follows. Section 2 explores a South African wind data set to monitor the wind behaviour over the course of a day. Section 3 outlines the WGSP and the PGSP models. Section 4 examines the behaviour of two distinct methodologies employed for evaluating the wind direction over multiple locations in South Africa. In Section 5, we will delve into the last remarks and potential avenues for future research.

2. Wind direction data in South Africa

The data utilised were obtained from the Council for Scientific and Industrial Research (CSIR) database³. Data from 97 locations which are relatively close to each other are considered for four different time periods (South African Standard Time (SAST)) on a particular day; 2012-12-31-05:00, 2012-12-31-11:00, 2012-12-31-17:00 and 2012-12-31-23:00. The region under consideration spans between 32.054° S, 24.009° E and 33.992° S, 27.99° E, which gives an area of about 214.908 km × 370.723 km which is roughly 79671.338 km². This region is illustrated in Figure 1. The data set included the wind direction in degrees, the longitude coordinate and the latitude coordinate.

The original wind direction is recorded in degrees, indicating the direction from which the wind originates. This is known as the meteorological wind direction (see Riha, 2020). Wind directions in degrees are converted to radians. Thus, a reading of 360° or 0°, which is equal to 0 rad or 2π rad, indicates wind coming from the north. Similarly, a reading of 90°, equivalent to $\pi/2$ rad, shows wind from the east, while 180° or π rad denotes wind from the south, and so on. The longitude and latitude coordinates are formatted in the Universal Transverse Mercator (UTM) format. Table 1 provides the circular descriptive statistics of the wind direction over the entire region under consideration for the four different time periods.

The dominant wind direction for the region under consideration is Northerly and North-Easterly at 05:00, North-Easterly at 11:00, Northerly at 17:00 and 23:00 as shown in the rose diagrams presented in Figure 2. Based on the descriptive measures in Table 1 and rose diagrams in Figure 2, it can be noted that the wind behaviour displays similar dominant directions for the two morning time periods (05:00 and 11:00) with a more North-Easterly pattern and the two evening time periods (17:00 and 23:00) with a Northerly pattern.

³<http://wasadata.csir.co.za/wasa1/WASAData> [Accessed July 2023]

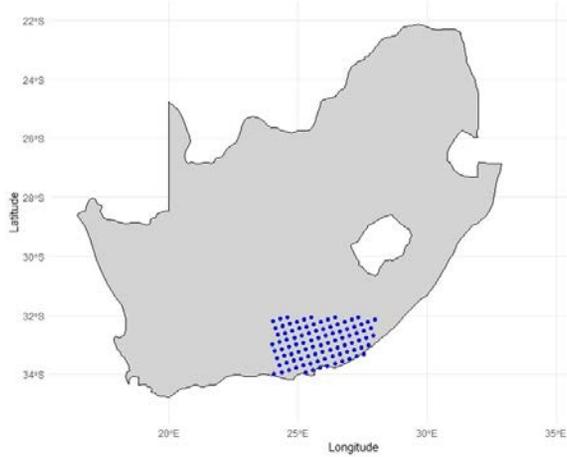


Figure 1. Map of South Africa with region under consideration indicated with dots.

Table 1. Circular descriptive statistics of the wind direction over the entire region under consideration for each time period.

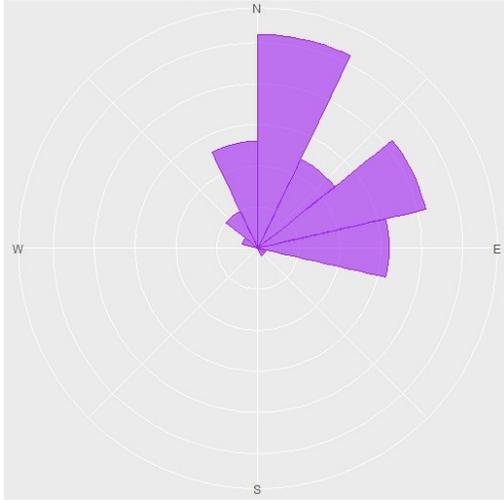
Time of day	mean direction	median direction	variance	standard deviation
05:00	0.62854	0.55833	0.23251	0.72750
11:00	0.51324	0.52081	0.10676	0.47517
17:00	0.14884	0.11990	0.05948	0.35019
23:00	0.20445	0.18064	0.07174	0.38585

3. Methodology

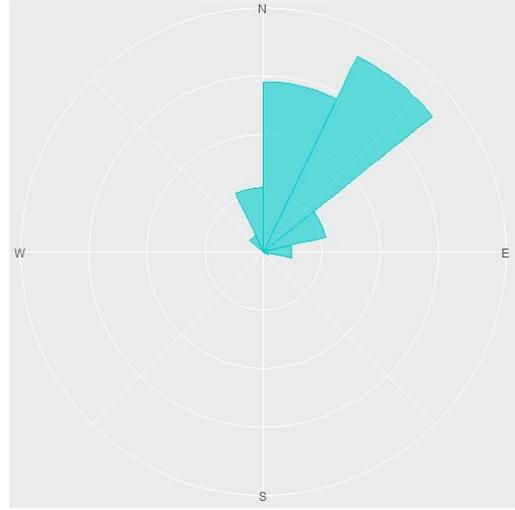
3.1 Wrapped Gaussian Spatial Process

In the linear domain, suppose we define a multivariate distribution for $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$ with $\mathbf{Y} \sim g(\cdot)$, where $g(\cdot)$ is a p -variate distribution on \mathbb{R}^p indexed by ω ; a sensible choice for $g(\cdot)$ would be a p -variate Gaussian distribution. Let $\mathbf{K} = (K_1, K_2, \dots, K_p)$ be such that $\mathbf{Y} = \mathbf{X} + 2\pi\mathbf{K}$. Then $\mathbf{X} = (X_1, X_2, \dots, X_p)$ is defined as a wrapped multivariate distribution induced from \mathbf{Y} with the transformation $\mathbf{X} = \mathbf{Y} \bmod 2\pi$. If the linear variable \mathbf{Y} is defined on \mathbb{R}^p then the wrapped induced variable \mathbf{X} will also be defined on \mathbb{R}^p as defined in Jupp and Mardia (2009). The wrapped Gaussian process will be fitted within a Bayesian framework using Markov Chain Monte Carlo (MCMC) methods. For further details the reader is referred to Jona-Lasinio et al. (2012).

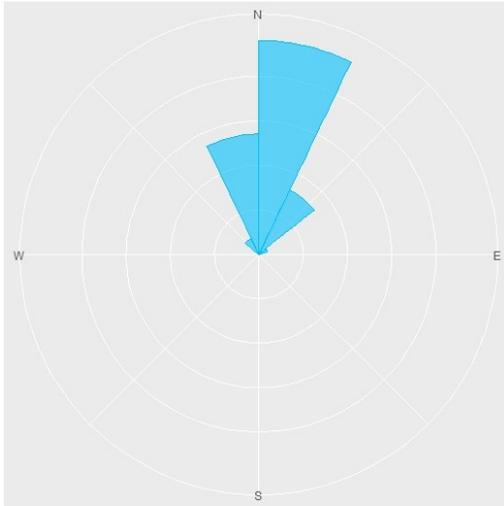
For the interpolation step, kriging will be used to make predictions at unobserved locations. Consider a Gaussian process (GP) in a spatial setting, we have locations s_1, s_2, \dots, s_p where $s \in \mathbb{R}^d$ and $Y(s)$ is a GP with mean $\mu(s)$ and an exponential covariance function $\sigma^2 \rho(s - s'; \phi)$ where ϕ is known as the decay parameter. We then have that $\mathbf{X} = (X(s_1), X(s_2), \dots, X(s_p))$ follows a wrapped Gaussian distribution with parameters $\boldsymbol{\mu} = (\mu(s_1), \dots, \mu(s_p))$ and $\sigma^2 \mathbf{R}(\phi)$ where $R(\phi)_{ij} = \rho(s_i - s_j; \phi)$ as defined in Jona-Lasinio et al. (2012). Suppose we have observations, $\mathbf{X} = (X(s_1), X(s_2), \dots, X(s_p))$, and would like to predict a new value $X(s_0)$ at an unobserved location s_0 . The point of departure follows similarly to a GP in the inline (linear) domain. The joint distribution for the linear observations



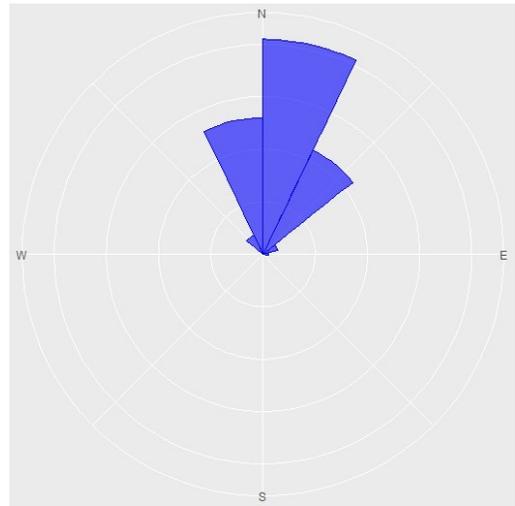
(a) Rose diagram of wind direction over the entire region at 05:00.



(b) Rose diagram of wind direction over the entire region at 11:00.



(c) Rose diagram of wind direction over the entire region at 17:00.



(d) Rose diagram of wind direction over the entire region at 23:00.

Figure 2. Rose diagram of wind direction over the entire region for the four different time periods.

$\mathbf{Y} = (Y(s_1), Y(s_2), \dots, Y(s_p))$ along with the unobserved $Y(s_0)$ is given as

$$\begin{bmatrix} \mathbf{Y} \\ Y(s_0) \end{bmatrix} \sim \mathbf{N} \left(\begin{bmatrix} \boldsymbol{\mu} \\ \mu(s_0) \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{R}_{\mathbf{Y}}(\phi) & \boldsymbol{\rho}_{0,\mathbf{Y}}(\phi) \\ \boldsymbol{\rho}_{0,\mathbf{Y}}^T(\phi) & 1 \end{bmatrix} \right). \quad (1)$$

From (1), the conditional distribution of $Y(s_0)|\mathbf{Y}, \boldsymbol{\theta}$ can be obtained. The wrapped Gaussian distribution of $X(s_0)|\mathbf{X}, \mathbf{K}, \boldsymbol{\theta}$, and thus $E(e^{iX(s_0)}|\mathbf{X}, \mathbf{K}; \boldsymbol{\theta})$, can then easily be derived. To obtain $E(e^{iX(s_0)}|\mathbf{X}, \mathbf{K}; \boldsymbol{\theta})$ it is necessary to marginalise over the distribution of $\mathbf{K}|\mathbf{X}, \boldsymbol{\theta}$ which will require

an n -fold sum over a multivariate discrete distribution which is problematic even when considering truncation. Thus, we consider a Bayesian framework to fit the wrapped GP model which will induce posterior samples $(\theta_b^*, \mathbf{K}_b^*)$, $b = 1, 2, \dots, B$. Using Monte Carlo integration the following approximation is obtained:

$$E(e^{i\mathbf{X}(s_0)} | \mathbf{X}) \approx \frac{1}{B} \sum_b \exp(-\sigma^2(s_0, \theta_b^*)/2 + i\tilde{\mu}(s_0, \mathbf{X} + 2\pi\mathbf{K}_b^*; \theta_b^*)). \quad (2)$$

The posterior mean kriged direction is

$$\mu(s_0, \mathbf{X}) = \arctan^*(g_{0,s}(\mathbf{X}), g_{0,c}(\mathbf{X})), \quad (3)$$

and the posterior kriged concentration is

$$c(s_0, \mathbf{X}) = \sqrt{(g_c(s_0, \mathbf{X}))^2 + (g_s(s_0, \mathbf{X}))^2}, \quad (4)$$

which is induced if $g_c(s_0, \mathbf{X}) = B^{-1} \sum_{b^*} \exp(-\sigma^2(s_0, \theta_b^*)/2) \cos(\tilde{\mu}(s_0, \mathbf{X} + 2\pi\mathbf{K}_b^*; \theta_b^*))$ and $g_s(s_0, \mathbf{X}) = B^{-1} \sum_{b^*} \exp(-\sigma^2(s_0, \theta_b^*)/2) \sin(\tilde{\mu}(s_0, \mathbf{X} + 2\pi\mathbf{K}_b^*; \theta_b^*))$.

3.2 Projected Gaussian Spatial Process

Suppose a random vector $\mathbf{Y} = (Y_1, \dots, Y_p)'$ follows a p -dimensional multivariate Gaussian distribution, with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ ($p \geq 2$). Then the unit vector $\mathbf{U} = \mathbf{Y}/\|\mathbf{Y}\|$ follows a projected Gaussian distribution with the same parameters and is denoted as $PN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as defined in Jupp and Mardia (2009). When $p = 2$, we obtain the circular projected Gaussian distribution. By projecting a bivariate spatial process on \mathbb{R}^2 , we can construct a spatial stochastic process of random variables taking values on a circle. Letting $(\cos \mathbf{X}(s), \sin \mathbf{X}(s))' = (Y_1(s), Y_2(s))'/\|\mathbf{Y}(s)\|$, we obtain the circular process $\mathbf{X}(s)$. This projected process inherits properties of the inline (linear) bivariate process such as stationarity. If we let $\mathbf{Y}(s)$ be a bivariate GP with mean $\boldsymbol{\mu}(s)$ and cross-covariance function $C(s, s') = \text{cov}(\mathbf{Y}(s), \mathbf{Y}(s'))$, then the induced circular process upon projection is defined as the projected Gaussian spatial process (PGSP). For the choice of the cross-covariance function we let $C(s, s') = \zeta(s, s') \cdot T$ where ζ is a valid correlation function and

$$T = \begin{pmatrix} \tau^2 & \rho\tau \\ \rho\tau & 1 \end{pmatrix}$$

is a 2×2 positive definite matrix as defined in Ley and Verdebout (2018).

Similarly to the WGSP, a Bayesian modelling framework is proposed for kriging due to the complexity of the conditional distributions of the GP. For the Bayesian formulation, we consider a conjugate prior, the bivariate Gaussian prior for $\boldsymbol{\mu}$. For τ^2 an inverse gamma with mean 1 is considered, and for ρ a uniform $(-1, 1)$ prior. For the decay parameter ϕ of the exponential correlation function, a uniform prior with support allowing ranges larger than the maximum distance over the region is utilised. The reader is referred to Ley and Verdebout (2018) for further details related to the projected Gaussian process.

4. Results and Discussion

The R (version 4.2.3 (2023-03-15 ucrt), R Core Team, 2023) software package **CircSpaceTime** developed by Jona-Lasinio et al. (2020) was used for the modelling of the wind direction data in South

Africa, specifically the **WrapSp**, **ProjSp**, **WrapKrigSp** and **ProjKrigSp** functions. **CircSpaceTime** was specifically developed for the implementation of Bayesian models for spatial interpolation of directional data using the wrapped Gaussian distribution and the projected Gaussian distribution.

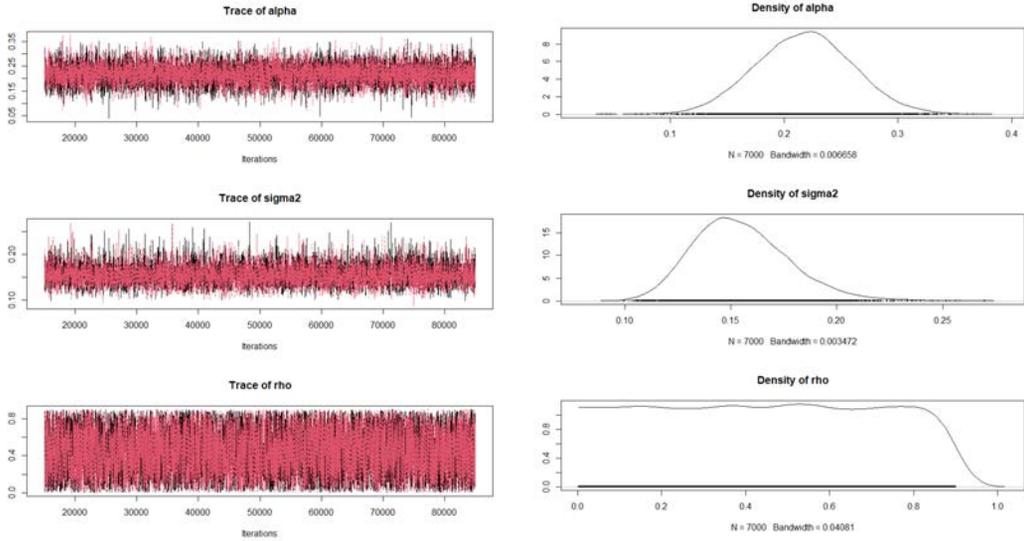
Firstly, the **WrapSp** function was applied to estimate the wrapped Gaussian posterior distribution for the given wind data. The **WrapSp** function can run for multiple MCMC chains, storing the posterior samples for μ (circular mean), σ^2 (variance) and ϕ (spatial correlation decay parameter). Based on the data described in Section 2, there were 97 observations ($n = 97$), 87 of which were used for the modelling, while the other 10 observations were our validation set. The validation set, consisting of 10 randomly selected points from the 97 observations, were used for prediction and model diagnostics. The **WrapSp** function requires the specification of prior distributions and a few parameters for the MCMC computation. The prior distribution values were chosen based on the data exploration as discussed in Section 2 and Table 1.

An exponential covariance function was considered. The prior for μ was a wrapped Gaussian distribution, for σ^2 an informative inverse gamma prior, and for the decay parameter ϕ a uniform prior which is weakly informative. The details of the model specification was provided for the 23:00 time period only. The remaining time periods follow similarly. Therefore, the prior distribution values applied for the 23:00 time period data were

- $\mu \sim \text{WN}(0, 2)$,
- $\sigma^2 \sim \text{IG}(7, 0.5)$,
- $\phi \sim \text{U}(0.001, 0.9)$.

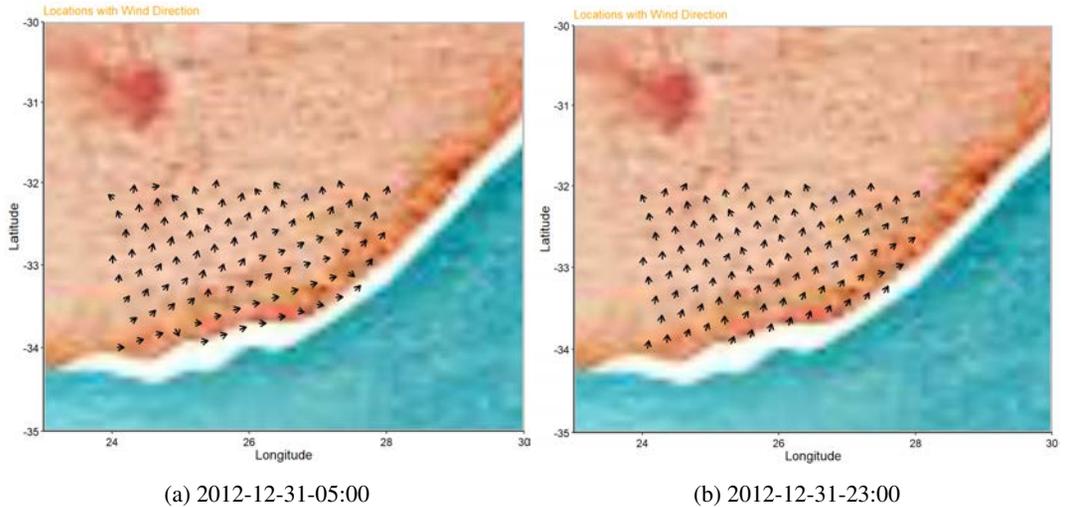
The MCMC ran with two chains in parallel for 100 000 iterations with a burnin of 30 000, thinning of 10 and an acceptance probability of 0.234 following Jona-Lasinio et al. (2012). The adaptive process of the Metropolis-Hasting step starts at the 100th iteration and ends at the 10 000th iteration. It is important that the adaptive procedure ends before the burnin is initiated to guarantee that the saved samples were drawn from correct posterior distributions as in Jona-Lasinio et al. (2020). The **ConvCheck** function was used to check for convergence and to obtain graphs of the MCMC. Figure 3 illustrates the traces and densities of the MCMC. A traceplot is an essential plot for evaluating convergence and diagnosing chain problems. It shows the time series of the sampling process and the expected outcome is to get a traceplot that looks completely random. The traceplots and the estimated posterior density plots of the generated samples are shown in Figure 3 for each of the parameters. Using our fitted model, **WrapKrigSp** was applied for the interpolation. The function produces posterior spatial predictions on the unobserved locations across all posterior samples, together with the mean and variance of the corresponding linear Gaussian process. Once the predictions were obtained, the average prediction error (APE) – defined as the average circular distance – and circular continuous ranked probability score (CRPS) were computed for the model; see Jona-Lasinio et al. (2012) and Jona-Lasinio et al. (2020).

From Table 2 we observe 95% credible intervals for $\hat{\mu}$, σ^2 , ϕ as well as the APE and CRPS for the wrapped Gaussian model. One must take into account that the $\hat{\mu}$ is a directional variable. The APE scores can be attributed to the fact that only 97 observations were considered. The APE score demonstrates sensitivity to the selection of prediction points, resulting in variability when different coordinates were used in the validation set. The APE is very dependent on the number of observations



(a) Trace plots of MCMC run for μ (top), σ^2 (middle) and ϕ (bottom). (b) Density plots of MCMC run for μ (top), σ^2 (middle) and ϕ (bottom).

Figure 3. Traces and densities from the MCMC run for the wrapped Gaussian spatial model.



(a) 2012-12-31-05:00 (b) 2012-12-31-23:00

Figure 4. South Africa: observed wind directions over the considered region at different time periods.

considered and the prior selection for ϕ . These results align with the conclusion in Riha (2020), who emphasises the importance of hyper-parameter settings for the prior distributions of the spatial decay parameter ϕ and the variance σ^2 for spatial interpolation with wrapped Gaussian process models. We note that the APE is affected by the data's variability. As depicted in Figure 4 and observed in Table

Table 2. The 95% credible intervals for $\hat{\mu}$, σ^2 , ϕ , the APE and CRPS for the WGSP model for the different time periods.

Time of day	$\hat{\mu}$ 95% C.I.	σ^2 95% C.I.	ϕ 95% C.I.	APE	CRPS
05:00	(0.45406; 0.76958)	(0.40963; 0.74301)	(0.02375; 0.87666)	0.66563	0.45200
11:00	(0.41816; 0.62425)	(0.17859; 0.31959)	(0.01624; 0.58459)	0.58396	0.47646
17:00	(0.07741; 0.22503)	(0.09163; 0.15986)	(0.01686; 0.58482)	0.12439	0.06647
23:00	(0.13639; 0.30191)	(0.11557; 0.20323)	(0.02382; 0.87776)	0.21914	0.14669

1 and 2, there is a noticeable contrast in data variance between the time periods 05:00 and 23:00. Specifically, at 05:00, wind directions exhibit significant variability, whereas at 23:00, they tend to align in a more consistent direction. Consequently, this disparity in data variability contributes to the difference in APE scores between these time periods. A similar pattern emerges when comparing the conditions at 11:00 and 17:00. It can be noted that the two morning time periods have much more variability than the two evening time points, with 17:00 having the lowest variance of 0.05948 yielding the lowest APE of 0.12439 as well.

Next we fit the PGSP model to the wind data observed at the 23:00 time period. Note the PGSP is more sensitive to the choice of priors, specifically for the decay parameter. The details of the model specification were provided for the 23:00 time period only. The remaining time periods follow similarly. The prior distribution values used in the PGSP model were

- $\mu \sim N\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}\right)$,
- $\sigma^2 \sim \text{IG}(7, 0.5)$,
- $\phi \sim \text{U}(0.001, 0.9)$,
- $\tau \sim \text{U}(-1, 1)$.

We specify an exponential covariance function to be used. The prior for μ was a bivariate Gaussian distribution, for σ^2 an informative inverse gamma prior, for τ a uniform prior and for the decay parameter ϕ a uniform prior which is weakly informative. The remainder of the function specification was the same as the WGSP model. From the convergence check (including the traceplots that were not reported) we see that the chains reached convergence. The PGSP's flexibility allows a better fit of the model with an APE of 0.03874 and a CRPS of 0.02506 for the 23:00 time period.

Table 3 reports the results of the APE and CRPS for both the WGSP and PGSP models. It is clear that for the South African wind data the PGSP model outperforms and is able to better capture the structure of the data. The WGSP model was computationally less demanding and allowed the choice of less informative priors, as well as for the parameters to be easily interpretable, which is not the case for the projected Gaussian model.

5. Conclusion

This paper explored the potential of utilising directional statistics within spatial analysis to model wind patterns in South Africa, drawing on methods developed in Jona-Lasinio et al. (2012). The

Table 3. Goodness-of-fit measures for the wrapped Gaussian model (WGSP) and projected Gaussian model (PGSP) for the South African wind data.

Time of day	Model	APE	CRPS
05:00	WGSP	0.66563	0.45200
	PGSP	0.12156	0.09955
11:00	WGSP	0.58396	0.47646
	PGSP	0.06419	0.04361
17:00	WGSP	0.12439	0.06647
	PGSP	0.04839	0.04319
23:00	WGSP	0.21914	0.14669
	PGSP	0.03874	0.02506

wrapped Gaussian model and projected Gaussian model were considered to account for the cyclic nature of the wind directions while also accounting for the spatial dependence. Based on the APE and CRPS, we conclude that the projected Gaussian process is an effective and precise approach to modelling wind patterns in South Africa. The model can adeptly manage directional data indexed by space, capturing the spatial structure among these observations. Looking ahead, enhancements to this model can be made through a more refined selection of parameters, like prior distributions, and by incorporating a more extensive set of locations to represent a broader area. Additionally, there is potential to expand this model into a spatio-temporal model, accounting for time as well. Another avenue for future work resides in accounting for the wind speed (and other wind characteristics) along with the wind directions.

In closing, the application of directional Gaussian processes in tandem with the capabilities of the **CircSpaceTime** package in R presents a compelling avenue for enhancing the accuracy and reliability of wind direction modelling. As the world increasingly recognises the critical role of sustainable energy sources, such as wind power, refining our understanding of wind behaviour becomes paramount, especially in South Africa with our current electricity problem. Better and more accurate understanding of wind behaviour can improve the design and optimisation of wind farms, thus ensuring efficient and effective harnessing of wind energy.

Acknowledgements. The authors would like to thank the anonymous reviewers for their insightful comments which led to an improvement of this paper. This work was based upon research supported in part by the National Research Foundation (NRF) of South Africa (Grant SRUG2204203965), as well as DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS). This project falls under the ethics application NAS116/2019.

References

- ARASHI, M., NAGAR, P., AND BEKKER, A. (2020). Joint probabilistic modeling of wind speed and wind direction for wind energy analysis: A case study in Humansdorp and Noupoort. *Sustainability*, **12**, 4371.
- BAO, L., GNEITING, T., GRIMIT, E. P., GUTTORP, P., AND RAFTERY, A. E. (2010). Bias correction

- and Bayesian model averaging for ensemble forecasts of surface wind direction. *Monthly Weather Review*, **138**, 1811–1821.
- CASSON, E. AND COLES, S. (1998). Extreme hurricane wind speeds: estimation, extrapolation and spatial smoothing. *Journal of Wind Engineering and Industrial Aerodynamics*, **74**, 131–140.
- COLES, S. (1998). Inference for circular distributions and processes. *Statistics and Computing*, **8**, 105–113.
- ENGEL, C. AND EBERT, E. (2007). Performance of hourly operational consensus forecasts (OCFs) in the Australian region. *Weather and Forecasting*, **22**, 1345–1359.
- JAMMALAMADAKA, S. R. AND SENGUPTA, A. (2001). *Topics in Circular Statistics*, volume 5. World Scientific.
- JONA-LASINIO, G., GELFAND, A., AND JONA-LASINIO, M. (2012). Spatial analysis of wave direction data using wrapped gaussian processes. *The Annals of Applied Statistics*, 1478–1498.
- JONA-LASINIO, G., ORASI, A., DIVINO, F., AND CONTI, P. L. (2007). Statistical contributions to the analysis of environmental risks along the coastline. *Società Italiana di Statistica-rischio e previsione. Venezia*, 6–8.
- JONA-LASINIO, G., SANTORO, M., AND MASTRANTONIO, G. (2020). CircSpaceTime: An R package for spatial and spatio-temporal modelling of circular data. *Journal of Statistical Computation and Simulation*, **90**, 1315–1345.
- JUPP, P. E. AND MARDIA, K. V. (2009). *Directional Statistics*. John Wiley & Sons.
- KALNAY, E. (2002). Atmospheric modeling, data assimilation and predictability.
- KATO, S., SHIMIZU, K., AND SHIEH, G. S. (2008). A circular-circular regression model. *Statistica Sinica*, 633–645.
- KENT, J. (1978). Limiting behaviour of the von Mises-Fisher distribution. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 84. Cambridge University Press, 531–536.
- LEY, C. AND VERDEBOUT, T. (2017). *Modern Directional Statistics*. CRC Press.
- LEY, C. AND VERDEBOUT, T. (2018). *Applied Directional Statistics: Modern Methods and Case Studies*. CRC Press.
- MARDIA, K. V. (1972). *Statistics of directional data Academic*. Elsevier.
- MASTRANTONIO, G., JONA-LASINIO, G., AND GELFAND, A. E. (2016). Spatio-temporal circular models with non-separable covariance structure. *Test*, **25**, 331–350.
- R CORE TEAM (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL: <https://www.R-project.org/>
- RAD, N. N., BEKKER, A., AND ARASHI, M. (2022). Enhancing wind direction prediction of South Africa wind energy hotspots with Bayesian mixture modeling. *Scientific Reports*, **12**, 11442.
- RIHA, A. E. (2020). *Hyperpriorsensitivity of Bayesian Wrapped Gaussian Processes with an Application to Wind Data*. Master’s thesis, Humboldt-Universität zu Berlin.
- STEPHENS, M. A. (1963). Random walk on a circle. *Biometrika*, **50**, 385–390.
- TRINH, V. AND CHUNG, C. (2023). Renewable energy for SDG-7 and sustainable electrical production, integration, industrial application, and globalization. *Cleaner Engineering and Technology*, **15**,

100657.

WATSON, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika*, **48**, 109–114.

WILKS, D. S. (2006). Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteorological Applications*, **13**, 243–256.

Information transmission between Bitcoin and other asset classes on the Johannesburg Stock Exchange

Kirsty Els, Caroline Mills, Wesley Turkington and Chun-Sung Huang

Department of Finance and Tax, University of Cape Town, Cape Town, South Africa

In this paper, we ascertain the information transmission and spillover effects between Bitcoin and the different asset classes or market sectors on the Johannesburg Stock Exchange. Specifically, our empirical investigation is conducted over the recent COVID-19 pandemic, the first major financial crisis on a global scale since the inception of Bitcoin. Our findings provide a pragmatic base of which investors, faced with changing market conditions, can make insightful investment decisions regarding hedging and diversification strategies through utilising cryptocurrencies in their existing portfolios. Cryptocurrencies, such as Bitcoin, may be argued as a new category of investment assets, and furthering our understanding of their behaviours may also assist policy makers and regulators of emerging economies. Using the multivariate VAR-BEKK-GARCH model, we are able to identify the directions and levels of volatility and return shock transmissions between Bitcoin and each asset class or market sector, ultimately determining the degree of integration between them. Moreover, conditional correlations between each pair may be extracted from the model and utilised to determine possible hedging and safe-haven opportunities through further analysis.

Keywords: Cryptocurrency, Information Transmission, Johannesburg Stock Exchange, Multivariate BEKK-GARCH, Volatility spillover.

1. Introduction

The infamous 2008 global financial crisis provoked low confidence among investors in the traditional centralised financial system. This ultimately catapulted the demand for cryptocurrencies, which are not subject to a governing body or interferences from a central bank. In addition, cryptocurrencies exhibit low trading costs and all transactions remain fully anonymous (Coeckelbergh and Reijers, 2016), which renders them suitable as an alternative currency or investment asset. Since the inception of Bitcoin as the first cryptocurrency in 2009, more than 22,000 different types of cryptocurrency have been created. In a short span of 14 years, the market capitalisation of cryptocurrencies has grown to an astonishing \$1.19 trillion in comparison to the global equity market of \$107 trillion (Maheshwari, 2023). This extraordinary growth has also drawn significant attention from both practitioners and academics alike. It has also become increasingly crucial to grasp the behaviours of cryptocurrencies, as well as their level of integration with other assets, which may help provide

Corresponding author: Chun-Sung Huang (chun-sung.huang@uct.ac.za)
MSC2020 subject classifications: 62P20, 91G70

regulatory bodies and policy makers with adequate guidance on cryptocurrencies as an investment tool (Vardar and Aydogan, 2019).

Speculators and investors have been attracted to cryptocurrencies due to their abnormal returns and high volatility levels. With reference to Bitcoin, the largest cryptocurrency by market capitalisation, the average volatility level since 2010 is at 114%, almost 10 times the volatility realised by typical equities and commodities, while obtaining annual returns reaching as high as approximately 254% (Blokland, 2021). These statistics indicate the high risk-reward characteristic of cryptocurrencies. Moreover, empirical evidence reveals that cryptocurrencies' high Sharpe Ratio, accompanied by low correlation with traditional asset classes, creates the potential for sizeable diversification and hedging benefits from holding cryptocurrencies in a traditional investment portfolio (Blokland, 2021).

Existing literature has predominantly focused on the characteristics of cryptocurrencies and how they compare to other asset classes, such as equities, foreign exchange and commodities (see, among others, Dyhrberg, 2016a; Pieters and Vivanco, 2017; Polasik et al., 2015; Yermack, 2015). On the other hand, the work of O'Dwyer (2015) primarily focused on the capacity of cryptocurrencies to create an alternative monetary system due to its characteristics of a more efficient, cheaper, and unregulated market space to transfer money (Vardar and Aydogan, 2019). However, a plethora of other studies from authors such as Wu et al. (2014) also advocated for cryptocurrencies to be considered as a completely new asset class that is independent of the behaviours of a traditional currency.

The motivations behind regarding cryptocurrencies, such as Bitcoin, as an alternative asset class instead of a traditional currency, is mainly premised on the discovery of typical stylised facts embedded within their empirical price returns data. For instance, evidence of leptokurtic behaviour was presented by Chan et al. (2017). Subsequently, the presence of heteroscedasticity and long memory properties were identified in the works of Gkillas and Katsiampa (2018) and Phillip et al. (2019), respectively. Such findings also further advocates for the use of GARCH-type models to estimate Bitcoin volatility (see, among others, Bouoiyour et al., 2016; Bouri et al., 2017; Dyhrberg, 2016a,b).

Notably, prior literature also provided evidence of low correlations between Bitcoin and other major financial asset classes (Baur et al., 2018). Such a phenomenon prompted overwhelming interest in utilising Bitcoin as a potential diversification and hedging tool to manage financial risks within existing investment portfolios (see, for instance, Briere et al., 2015; Dyhrberg, 2016b; Aslanidis et al., 2019; Fakhfekh and Jeribi, 2020). The evidence suggests that adding a small portion of cryptocurrencies to a diversified portfolio, made up of traditional assets, can substantially reduce the overall risk for a given level of expected return.

In addition to the above, practitioners and academics alike have been interested in the ability of cryptocurrencies to act as safe-havens during periods of market distress. Klein et al. (2018) made a valuable contributions to existing literature in this regard. Using the celebrated BEKK-GARCH model, the authors demonstrated that gold plays an important role in financial markets with 'flight-to-quality' in times of market distress. This is somewhat similar to Bitcoin as the cryptocurrency's returns are negatively correlated to downward moving markets (see, for instance, Klein et al., 2018). Such work has also led to further investigations regarding the ability of cryptocurrencies to act as a hedging strategy under different market conditions. Given the growing acceptance of cryptocurrencies, their information transmission with traditional financial markets is becoming increasingly

important for modern investors.

The recent outbreak of the infamous COVID-19 pandemic has led to the first major widespread global economic distress period since the inception of cryptocurrencies. This unprecedented opportunity, while cataclysmic, lends itself for researchers to more sufficiently analyse the capacity of cryptocurrencies to act as an effective hedging or safe-haven tool during periods of market distress on a global scale. However, recent studies have provided conflicting evidence regarding the ability of cryptocurrencies to act as an adequate safe-haven tool (see, for examples, Conlon et al., 2020; Raheem, 2021; Rubbaniy et al., 2021; Marobhe, 2022; Melki and Nefzi, 2022; Abdelmalek and Benlagha, 2023). Interestingly, contradictions on the above hedging and safe-haven characteristics were found both across different financial markets, as well as across different cryptocurrencies.

In this paper, we contribute to the inconclusive debate by investigating the extent of volatility and return spillovers between Bitcoin and the various asset classes or market sectors of the South African financial market. Over and above determining whether Bitcoin may indeed provide benefits of a safe-haven for South African investors, our empirical work adds to the discussion regarding whether cryptocurrencies should be considered as an independent asset class of its own.

The South African market has potential to produce valuable results when analysing the information transmission between cryptocurrencies and traditional financial assets. This can be attributed to the fact that it has one of the highest GDP per capita in Africa and a rapidly increasing internet penetration rate (Vincent and Evans, 2019). More importantly, in recent years, South Africa has been experiencing poor economic growth and political instability, prompting investors to look towards other alternatives, such as cryptocurrencies, that do not rely on governing authorities and central monetary systems. Noticeably, South Africa is also considered one of the leading Bitcoin economies on the African continent (Vincent and Evans, 2019).

We conducted our investigation by employing a multivariate vector autoregressive (VAR) in mean GARCH framework with BEKK representations, as proposed by Engle and Kroner (1995). The model provides a valuable opportunity for us to interrogate the information spillover effects in both the return and volatility between Bitcoin and other asset classes or market sectors on the Johannesburg Stock Exchange (JSE). More importantly, our results lead to a better understanding of cryptocurrencies' ability to act as a possible hedging or safe-haven tool against traditional assets in South Africa. It is also worthwhile highlighting that our proposed work closely follows the methodology of Vardar and Aydogan (2019).

The rest of the paper is organised as follows. Section 2 introduces the data and our proposed methodology. Our empirical findings and ensuing discussions are presented in Section 3. Finally, Section 4 concludes our work and provides suggestions for further research.

2. Data and Methodology

The daily prices of Bitcoin and levels of the South African financial indices, namely, Top 40 (TOP40), Resource 20 (RESI), Financial 15 (FINI), Industrial 25 (INDI), and All Bonds (ALBI) forms the dataset of this study. All data points were obtained from Bloomberg and exclude weekends and public holidays. Our data samples used for the proposed analysis spans from 1 July 2019 to 30 April 2021, which includes the periods immediately leading up to the formal declaration of the COVID-19 pandemic on 11 March 2020, as well as the subsequent period that ensued. Daily returns of each time

series were calculated with the usual natural logarithmic procedure, $r_t = \ln(S_t) - \ln(S_{t-1})$, where S_t is the spot price of the financial asset at time t .

While the ALBI is an adequate representative of the bonds asset class, the Top 40 is primarily used as a benchmark for equities. All listed entities on the JSE are categorised into one of the three sectors of Resources, Financials and Industrials as per the Industry Classification Benchmark (ICB), based on their revenue. In particular, RESI includes the largest 20 entities by market capitalisation and identified as Basic Materials and Energy, while FINI exhausts the list of the largest 15 entities by market capitalisation that are characterised as Financials and Real Estates. Lastly, the largest 25 entities by market capitalisation in the remaining pool not classified as above are absorbed by the INDI. This distinct separation on the JSE provides us with the opportunity to further examine the ability of Bitcoin to act as a possible hedging or safe-haven tool against different market sectors apart from just asset classes.

In this paper, we deploy the multivariate vector autoregressive GARCH framework with BEKK specifications (VAR-BEKK-GARCH), as proposed by Engle and Kroner (1995). An important feature of the VAR-BEKK-GARCH model is the absence of restrictions imposed on the correlation structure between the variables in question. Moreover, the BEKK specification has the advantage of allowing for information spillover to be observed from both directions of the time series pair in question.

The VAR specification in the conditional mean equations allow us to evaluate the spillover in mean. Through minimising the Akaike Information Criterion, we select the following VAR(1) model to represent the BEKK-GARCH-in-mean equation:

$$R_t = \mu + \Phi R_{t-1} + \epsilon_t, \quad (1)$$

where $R_t = (r_t^c, r_t^s)'$. We denote r_t^c and r_t^s as the logarithmic return of Bitcoin and the logarithmic return of a chosen financial index at time t , respectively. Specifically, the financial indices are either the TOP40, RESI, FINI, INDI and ALBI at time t . $\mu = (\mu_1, \mu_2)'$ is a vector of the constant terms of the conditional mean equations. The (2×2) matrix of coefficients for the lag variables in the VAR(1) mean specification is denoted by

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}.$$

Lastly, $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ is a vector of residuals for the cryptocurrency and a financial index, respectively, and are both assumed to be normally distributed with a mean of 0.

The conditional variance-covariance matrix (H_t) of the residuals is defined as follows:

$$\epsilon_t | \Omega_{t-1} \approx N(0, H_t), \quad H_t = \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix}, \quad (2)$$

where Ω_{t-1} is the set of all information up until time $t-1$. The conditional covariances, represented by h_{12} and h_{21} , captures the relationship between Bitcoin and a financial index in question. Specifically, the BEKK-GARCH(1,1) model can be expressed as

$$H_t = C'C + A'\epsilon_{t-1}\epsilon_{t-1}'A + B'H_{t-1}B, \quad (3)$$

where C is a (2×2) upper triangular matrix of constants for the cryptocurrency and stock index pair; A is the (2×2) matrix of ARCH coefficients that capture the effects of local and cross-market

shocks, while B the corresponding (2×2) matrix of GARCH coefficients that capture the effect of own market volatility persistence and the cross-market volatility transmissions, i.e. between Bitcoin and a financial index. Specifically, our binary BEKK-GARCH(1,1) model, as per expression (3), may be expanded as follows:

$$\begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}_t = \begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix}' \begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}' \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix}' \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \\ + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}' \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}_{t-1} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}. \quad (4)$$

We may further express (4) with the following set of equations:

$$\begin{aligned} h_{11,t} &= c_{11}^2 + a_{11}^2 \varepsilon_{1,t-1}^2 + 2a_{11}a_{21} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{21}^2 \varepsilon_{2,t-1}^2 + b_{11}^2 h_{11,t-1} + 2b_{11}b_{21} h_{21,t-1} \\ &\quad + b_{21}^2 h_{22,t-1}, \\ h_{12,t} &= c_{11}c_{12} + a_{11}a_{12} \varepsilon_{1,t-1}^2 + (a_{11}a_{22} + a_{12}a_{21}) \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{21}a_{22} \varepsilon_{2,t-1}^2 + b_{11}b_{12} h_{11,t-1} \\ &\quad + (b_{11}b_{22} + b_{12}b_{21}) h_{12,t-1} + b_{21}b_{22} h_{22,t-1}, \\ h_{22,t} &= c_{12}^2 + c_{22}^2 + a_{12}^2 \varepsilon_{1,t-1}^2 + 2a_{12}a_{22} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{22}^2 \varepsilon_{2,t-1}^2 + b_{12}^2 h_{11,t-1} + 2b_{12}b_{22} h_{21,t-1} \\ &\quad + b_{22}^2 h_{22,t-1}, \end{aligned} \quad (5)$$

where $h_{11,t}$ and $h_{22,t}$ are the conditional variances of Bitcoin and a financial index, respectively. Similarly, $h_{12,t}$ and $h_{21,t}$ represents the conditional covariances across the two respective assets. The VAR-BEKK-GARCH model parameters (μ, Φ, C, A, B) may be estimated using the quasi-maximum likelihood method, whereby the log-likelihood function for a sample of T observations is given by (Engle and Kroner, 1995)

$$\log L = -\frac{1}{2} \sum_{t=1}^T [k \log(2\pi) + \log |H_t| + \varepsilon_t' H_t^{-1} \varepsilon_t], \quad (6)$$

where L denotes the likelihood function used to estimate the vector of unknown model parameters, and k the number of variables ($k = 2$ for bi-variate form).

Equation set (5) also demonstrates that the conditional variance and covariances across the time series pair are not only influenced by the residuals of the two time series but also by the square of the residuals (i.e. $h_{11,t-1}$, $h_{12,t-1}$ and $h_{22,t-1}$). To determine the volatility spillover effects, we observe the resulting ARCH and GARCH effects, as well as the asymmetric effects of both positive and negative shocks. Specifically, when $a_{12} = b_{12} = 0$ the conditional variance of the chosen financial index is only affected by its own lagged squared residuals and lagged conditional variance, implying that Bitcoin has no volatility spillover effects on the chosen financial index. Similarly, $a_{21} = b_{21} = 0$ suggests that the chosen financial index has no volatility spillover effects on Bitcoin. Hence, utilising the significance of the coefficients from the VAR-BEKK-GARCH model, we may interrogate the mean and volatility spillover effects between Bitcoin and other financial sectors and asset classes.

Lastly, through conditional covariances of the VAR-BEKK-GARCH model, the dynamic correla-

tion between Bitcoin and other asset classes considered in this study can be obtained as follows:

$$\rho_t = \frac{h_{12,t}}{\sqrt{h_{11,t} \times h_{22,t}}}. \quad (7)$$

The dynamic conditional correlation (7) may be utilised to observe correlation fluctuations and its varying characteristics, lending itself as an useful risk measuring tool. Following Baur and Lucey (2010), we define assets that are uncorrelated (negatively correlated) with another asset or portfolio in periods of market crisis as a weak (strong) safe-haven tool.

3. Empirical Results and Discussions

3.1 Preliminary analysis

Prior to fitting the VAR-BEKK-GARCH model, we examine the graphical plots and descriptive statistics of each asset class and market sector returns, as well as perform conditional heteroscedasticity tests, to establish whether our dataset indeed satisfies the requirements of the VAR-BEKK-GARCH model.

Figure 1 illustrates the trend and variations in the daily log returns of Bitcoin and our chosen financial indices. Specifically, we can identify that Bitcoin returns display a different volatility pattern than the various other financial indices considered in this study. The graphical exhibition of our time series also suggests possible heteroscedasticity embedded in our returns data. Notably, Bitcoin also displays different periods of highs and lows in volatility relative to the other financial indices. The summary of descriptive statistics in Table 1 shows a significantly higher mean and standard deviation in Bitcoin returns in comparison to those of the various financial indices used in this study. Evidence from the skewness and excess kurtosis values, together with the rejections of the Jarque-Bera tests, clearly advocate for the absence of normality within our data. It should also be noted that all of our chosen time series are consistently characterised by a negatively-skewed distribution, which is commonly observed in financial data (Cont, 2001). The rejection of the Augmented Dickey-Fuller (ADF) test statistics demonstrates that our datasets are stationary series without unit root, suggesting suitability in assuming a VAR model. Finally, the rejection of the null in the Lagrange multiplier tests of Engle (1982) indicates the existence of heteroscedasticity effects, and supports our decision to estimate the series pair with a BEKK-GARCH model.

The linear Pearson correlation coefficients between Bitcoin and each financial index across our sample period is reported in Table 2. We observe from column 2 that Bitcoin exhibits a positive unconditional correlation with all the asset classes considered in this study, albeit distinctly weak or negligible with the FINI, INDI and ALBI. This suggests that Bitcoin may have the potential to act as a possible weak safe haven for bonds and entities on the JSE that are categorised as financials and industrials per the ICB. Our results are also consistent with prior findings of Bitcoin's positive relationship to gold and commodities (Wang et al., 2019).

3.2 Mean and volatility spillover effects

The empirical results of the VAR-BEKK-GARCH model to demonstrate mean and volatility spillovers between Bitcoin and the different financial indices (TOP40, RESI, FINI, INDI and ALBI) are presented in Table 3. Firstly, as per Panel A, the failure to reject the null hypothesis for every ϕ_{12}

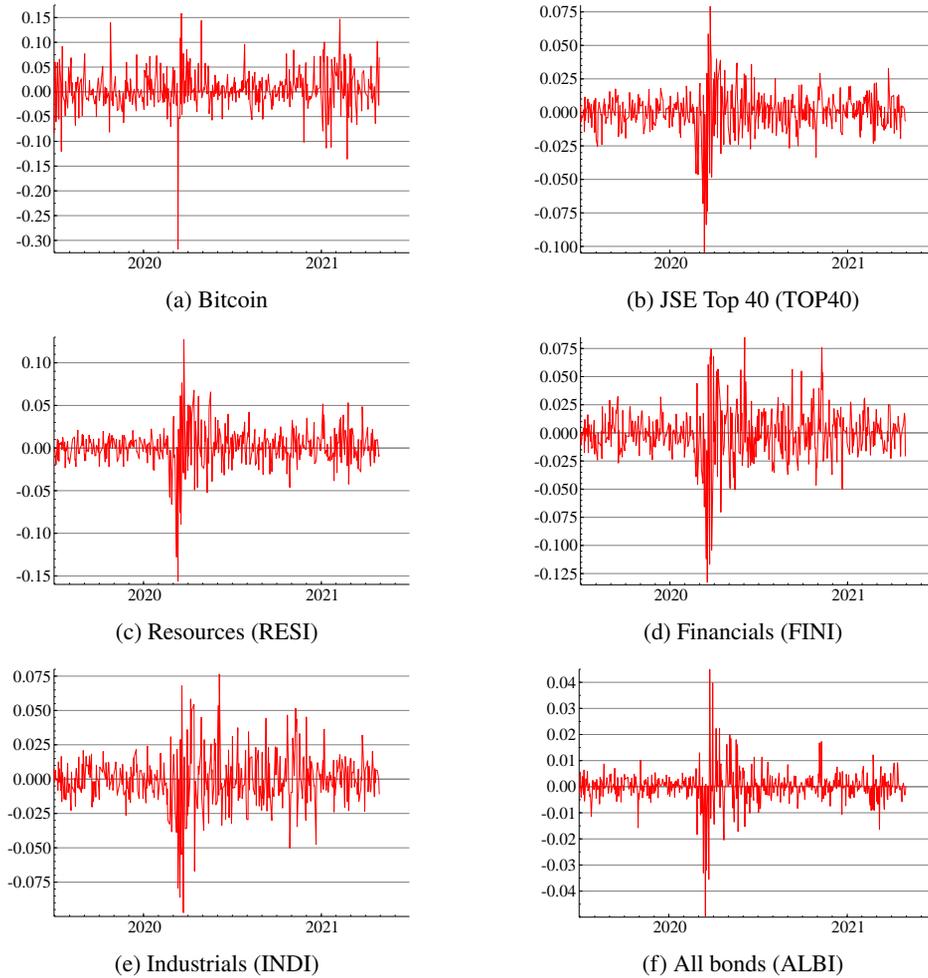


Figure 1. Plot of daily log returns.

clearly indicates an absence of mean spillover effects from each financial index into Bitcoin. However, the rejections of the null hypotheses across the various ϕ_{21} , except for RESI, show that Bitcoin tends to impose a mean spillover effect on the different financial indices at a 10% level of significance. Hence, all cross-mean effects are dominated by unilateral positive spillovers from Bitcoin to the different financial indices during the COVID-19 period. More concretely, the current-period returns across the various financial indices are influenced by the previous-period returns of Bitcoin. As a result, opportunities may exist for market participants to utilise Bitcoin returns to predict returns across the different asset classes and market sectors on the JSE.

From Panel B of Table 3, we use a_{12} , a_{21} , b_{12} and b_{21} to interrogate the shocks and volatility spillovers across Bitcoin and financial indices. Notably, we omit results from matrix C as it does not influence the volatility spillover effects. Failures in the rejection of the null for a_{12} and a_{21} for Bitcoin against the FINI, INDI and ALBI are empirical evidence to suggest an absence of shock spillover

Table 1. Descriptive statistics.

Variable	Bitcoin	TOP40	RESI	FINI	INDI	ALBI
Mean	0.379%	0.037%	0.079%	-0.068%	-0.030%	0.017%
SD	0.040	0.016	0.022	0.023	0.019	0.006
Min.	-0.317	-0.105	-0.156	-0.133	-0.097	-0.050
Max.	0.158	0.079	0.127	0.085	0.076	0.045
Skewness	-0.808	-1.024	-0.676	-0.687	-0.299	-0.478
Kurtosis	10.039	9.232	10.094	6.040	4.056	18.938
J-B	1986.8***	1718.5***	1993.2***	738.54***	324.2***	9583.5***
ADF	-8.098***	-7.774***	-6.854***	-7.936***	-8.245***	-8.517***
LM	22.267**	185.37***	146.8***	182.22***	109.06***	183.76***

Notes: SD is an abbreviation for standard deviation. J-B is an abbreviation for the Jarque-Bera static, which tests for the rejection of the null hypothesis of a normal distribution. ADF is a statistic of the Augmented Dickey-Fuller test for a unit root. LM is the Lagrange multiplier test of Engle (1982) for the detection of conditional heteroscedasticity. Lastly, ***, **, and * indicate that the null hypothesis is rejected at the 1%, 5%, and 10% significance levels, respectively.

Table 2. Unconditional correlation coefficient matrix.

	Bitcoin	TOP40	RESI	FINI	INDI	ALBI
Bitcoin	1					
TOP40	0.268	1				
RESI	0.267	0.878	1			
FINI	0.056	0.707	0.489	1		
INDI	0.094	0.644	0.484	0.802	1	
ALBI	0.087	0.423	0.298	0.523	0.440	1

(or ARCH) effects from Bitcoin to the three indices and vice versa. However, a unilateral shock spillover effect from TOP40 on Bitcoin was detected at the 5% level of significance. Additionally, the rejection of the null for both a_{12} and a_{21} between Bitcoin and RESI infers a significant bilateral shock spillovers between the pair.

The resulting values of b_{12} and b_{21} painted a similar pattern from the perspectives of volatility spillovers. A significant bilateral volatility transmission between Bitcoin and RESI was detected at the 1% level of significance. Similarly, there is an unilateral volatility spillover (or GARCH effect) from TOP40 into Bitcoin. However, an opposite unidirectional volatility spillover effect was found between Bitcoin and INDI at the 5% level of significance. Finally, there is a clear absence of volatility transmissions between Bitcoin and two indices, namely, FINI and ALBI, advocating the possibility of using Bitcoin as a safe-haven tool for entities categorised as financials or for South African bonds. Summaries of our directional spillover results are illustrated in Table 4.

3.3 Dynamic correlation

In addition to our spillover analysis, based on the VAR-BEKK-GARCH model and daily data, we obtain the dynamic conditional correlations between Bitcoin and each financial index as per equation (7). The descriptive statistics of the pairwise dynamic conditional correlations are reported in Table 5. All dynamic conditional correlations rejects the null hypothesis of normality, except for the Bitcoin and INDI pair. Plots of the dynamic conditional correlations are also illustrated in Figure 2. We observe that the correlation between Bitcoin and TOP40 exhibits the highest mean, indicating a

Table 3. VAR-BEKK-GARCH results.

	TOP40	RESI	FINI	INDI	ALBI
Panel A - mean equation					
ϕ_{11}	0.013368 [0.256]	0.010112 [0.200]	0.049376 [0.844]	0.031172 [0.684]	0.067321 [0.294]
ϕ_{12}	0.012378 [0.088]	0.12190 [1.188]	0.066726 [0.598]	-0.118099 [-1.297]	0.265320 [0.208]
μ_1	0.003597 [1.989]**	0.002929 [1.580]	0.003014 [1.686]*	0.003400 [1.918]*	0.003165 [1.641]
ϕ_{21}	0.028243 [1.885]*	0.015407 [0.538]	0.031543 [1.827]*	0.034186 [1.908]*	0.016550 [1.728]*
ϕ_{22}	0.031625 [0.589]	0.018887 [0.339]	0.096669 [1.577]	0.021106 [0.408]	0.159620 [1.478]
μ_2	0.000690 [1.372]	0.000356 [0.469]	0.000071 [0.118]	-0.000219 [-0.301]	0.000327 [1.762]*
Panel B - variance equation					
c_{11}	0.032502 [10.080]***	0.009823 [4.081]***	0.009646 [3.288]***	0.009933 [3.788]***	0.029935 [11.300]***
c_{12}	-0.001445 [-0.799]	-0.003282 [-2.771]***	-0.001195 [-0.860]	-0.002807 [-2.552]**	0.000911 [2.450]**
c_{22}	0.002437 [0.471]	0.000003 [1.091]	0.002092 [1.315]	0.000106 [1.698]*	0.000477 [0.202]
a_{11}	0.123158 [1.529]	0.078500 [0.564]	0.227533 [3.393]***	0.131688 [2.470]**	0.612290 [3.634]***
a_{12}	-0.018926 [-0.768]	-0.081901 [-2.175]**	-0.042484 [-1.241]	-0.022363 [-0.999]	-0.013805 [-0.336]
a_{21}	-0.460779 [-2.543]**	0.604144 [3.155]***	0.208485 [0.886]	0.053027 [0.6025]	2.107222 [1.592]
a_{22}	0.420980 [3.115]***	-0.109648 [-1.412]	0.386407 [4.855]***	-0.384917 [-6.969]***	0.294088 [2.225]**
b_{11}	0.220588 [1.808]*	0.867930 [27.290]***	0.941149 [35.410]***	0.942733 [29.990]***	0.236260 [3.095]***
b_{12}	-0.045780 [-0.519]	0.284597 [8.206]***	0.015834 [1.378]	0.147450 [2.313]**	-0.029659 [-0.337]
b_{21}	1.138973 [2.173]**	-0.998558 [-7.707]***	-0.046034 [-0.854]	0.233225 [0.8156]	-0.191528 [-0.237]
b_{22}	0.903661 [6.520]***	0.649709 [9.053]***	0.912537 [23.920]***	-0.884212 [-21.140]***	0.931632 [28.340]***

Notes: μ_1 and μ_2 are the constant terms of the respective mean equations of time series 1 (Bitcoin) and time series 2 (a financial index of either TOP40, RESI, FINI, INDI or ALBI), respectively. ϕ_{11} and ϕ_{22} represents each time series' own lagged mean effects, respectively. ϕ_{12} indicates the lagged spillover effects in mean from Bitcoin to the financial index in question, while ϕ_{21} is the same effect in the opposite direction. The constant terms of the variance equations are given by c_{11} , c_{22} and c_{21} . a_{11} and a_{22} are the ARCH effects in the two time series, respectively. Parameter a_{12} stands for the spillover effect from a prior shock in Bitcoin returns on the current volatility of a financial index, whereas a_{21} measures the same spillover effect in the opposite direction. b_{11} and b_{22} captures the GARCH effects that measures the persistence of volatility in time series 1 and 2, respectively. b_{12} represents the spillover effect of Bitcoin's variance in the previous time period to the current variance of the chosen financial index, while b_{21} shows the same spillover effect in the opposite direction. The corresponding t -statistics for the significance of the various parameters are presented in square parentheses, with ***, **, and * indicate that the null hypothesis is rejected at the 1%, 5%, and 10% significance levels, respectively.

stronger relationship between Bitcoin and TOP40 than other indices during a period of market crisis. This is consistent with our observations in Figure 2(a), where the dynamic conditional correlation between Bitcoin and TOP40 experienced a significant upward spike immediately following the COVID-19 outbreak. This further suggests Bitcoin to be an inadequate safe-haven tool for TOP40.

The mean conditional correlation between Bitcoin and RESI is relatively high. This can be expected due to the closeness in relationship between Bitcoin and exhaustible resource commodities, as well as precious metals, as advocated by an existing line of research (see, Gronwald, 2019; Mensi et al., 2019). Our empirical findings provide further evidence in support of such a phenomena, which stays persistent even during periods of market crisis such as the recent COVID-19 pandemic. This is also observable through our dynamic conditional correlation plot in Figure 2(b).

In line with our unconditional correlation in Table 2, lower means in conditional correlation are observed between Bitcoin and the 3 indices, namely, FINI, INDI and ALBI. However, as illustrated in Figure 2(d)(e), Bitcoin may be inadequate to serve as a safe-haven tool for both INDI and ALBI. Both pairs are affected by upticks and prolonged positive trends in the dynamic conditional correlation following the COVID-19 crisis. The findings against the ALBI are intriguing as it contradicts prior studies that motivated for Bitcoin to act as a hedging tool for bonds (see, Kang et al., 2020; Wang et al., 2019). Interestingly, with the lowest mean conditional correlation, we observe significant

Table 4. Directional summary of spillover results.

	TOP40	RESI	FINI	INDI	ALBI
Panel A - Mean Spillovers					
Bitcoin	→	-	→	→	→
Panel B - Shock Transmission					
Bitcoin	←	↔	-	-	-
Panel C - Volatility Spillovers					
Bitcoin	←	↔	-	→	-

Notes: The ↔ represents a bidirectional spillover, whereas → or ← indicates a unilateral transmission. We use - to show an absence of transmission. Specifically a → demonstrates that Bitcoin is a transmitter, while ← indicates that Bitcoin is a receiver.

Table 5. Descriptive statistics of dynamic correlations.

Variable	TOP40	RESI	FINI	INDI	ALBI
Mean	0.183	0.165	0.081	0.103	0.126
SD	0.165	0.155	0.127	0.181	0.178
Min.	-0.346	-0.552	-0.618	-0.400	-0.551
Max.	0.820	0.791	0.610	0.688	0.882
Skewness	0.777	-0.331	-1.073	0.054	-0.050
Kurtosis	2.063	2.222	5.133	0.076	1.787
J-B	128.58***	103.89***	594.39***	0.38	85.90***

Notes: SD is an abbreviation for standard deviation. J-B is an abbreviation for the Jarque-Bera statistic, which tests for the rejection of the null hypothesis of a normal distribution. Lastly, ***, **, and * indicate that the null hypothesis is rejected at the 1%, 5%, and 10% significance levels, respectively.

downward ticks and ensuing negative correlations between Bitcoin and FINI following the COVID-19 crisis. Moreover, the dynamic conditional correlation remained low even after reverting to positive trends, suggesting adequacy in Bitcoin to act as a possible strong safe-haven tool for JSE entities categorised as financials. Notably, the dynamic conditional correlation between the pair also exhibits the lowest standard deviation, indicating the least violent fluctuations in comparison to the movements in conditional correlation between Bitcoin and other indices.

4. Conclusion

The debate of whether cryptocurrencies may act as an adequate hedging or safe-haven tool for traditional financial assets remains a contentious one for academics and practitioners alike. In this paper, we further contributed to the literature by investigating the safe-haven characteristics of Bitcoin for some traditional asset classes on the Johannesburg Stock Exchange. Specifically, our empirical analysis is performed over the recent infamous period of COVID-19, an unprecedented period of major financial market turmoil since the inception of cryptocurrencies. We provide evidence to further demonstrate the close relationship between Bitcoin and commodities, as represented by the RESI index, and showed the consistency of such an interconnectedness between the pair even during periods of extreme market distress. In addition, our results contradicted the existing acceptance that Bitcoin is an adequate hedging tool for bonds. During periods of market crisis, bonds may

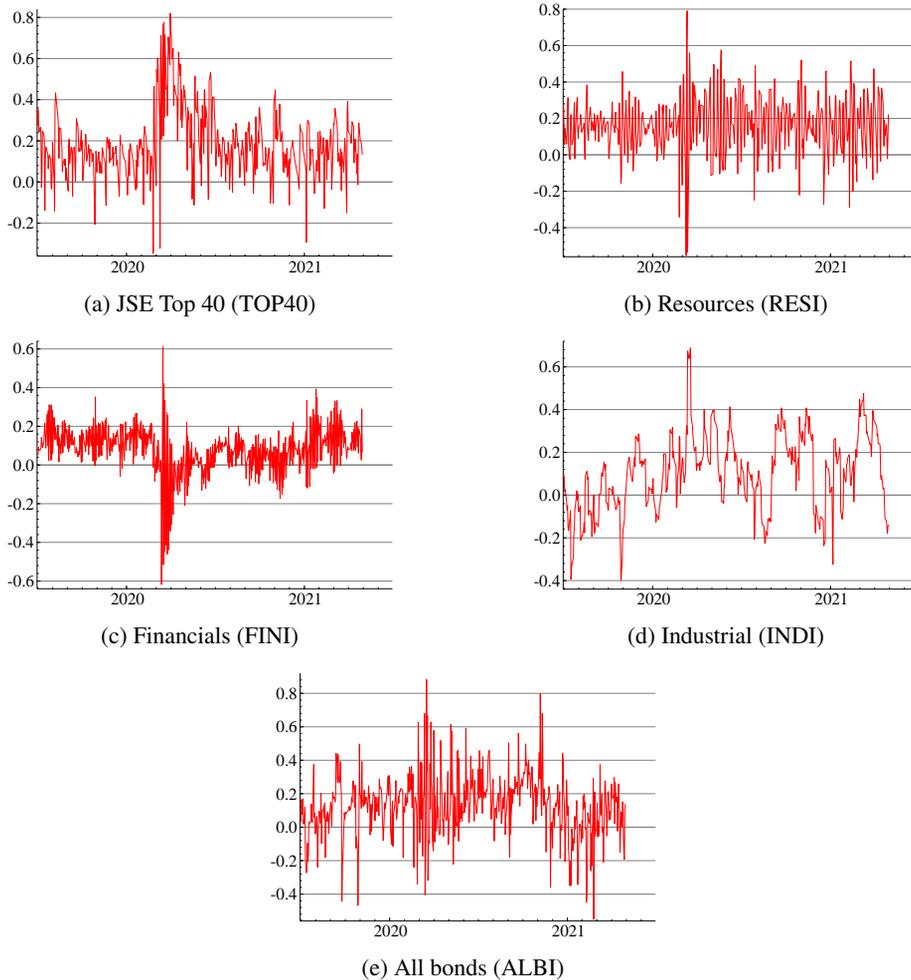


Figure 2. Plots of conditional correlations between Bitcoin and financial indices.

indeed be inadequate to act as a safe-haven tool. Finally, our empirical analysis showed that JSE entities classified as financial and real estate, as per the ICB, may turn to Bitcoin as a potential strong safe-haven tool. Limitations of our studies may be remedied by first including other widely traded cryptocurrencies that have already gained significant market capitalisation, and analyse the adequacy of these cryptocurrencies to act as potential safe havens for traditional asset classes during periods of market crisis. Moreover, in-depth investigations on the effect of extreme quantiles of traditional asset classes and its subsequent effect on their dynamic conditional correlations with different cryptocurrencies may add compelling evidence to the ongoing debate.

References

- ABDELMALEK, W. AND BENLAGHA, N. (2023). On the safe-haven and hedging properties of Bitcoin: New evidence from COVID-19 pandemic. *The Journal of Risk Finance*, **24**, 145–168.
- ASLANIDIS, N., BARIVIERA, A. F., AND MARTÍNEZ-IBAÑEZ, O. (2019). An analysis of cryptocurrencies conditional cross correlations. *Finance Research Letters*, **31**, 130–137.
- BAUR, D. G., HONG, K., AND LEE, A. D. (2018). Bitcoin: Medium of exchange or speculative assets? *Journal of International Financial Markets, Institutions and Money*, **54**, 177–189.
- BAUR, D. G. AND LUCEY, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Financial Review*, **45**, 217–229.
- BLOKLAND, J. (2021). Bitcoin as digital gold – a multi-asset perspective.
URL: <https://www.robeco.com/en-za/insights/2021/04/bitcoin-as-digital-gold-a-multi-asset-perspective>
- BOUOUIYOUR, J., SELMI, R., ET AL. (2016). Bitcoin: A beginning of a new phase. *Economics Bulletin*, **36**, 1430–1440.
- BOURI, E., JALKH, N., MOLNÁR, P., AND ROUBAUD, D. (2017). Bitcoin for energy commodities before and after the December 2013 crash: Diversifier, hedge or safe haven? *Applied Economics*, **49**, 5063–5073.
- BRIERE, M., OOSTERLINCK, K., AND SZAFARZ, A. (2015). Virtual currency, tangible return: Portfolio diversification with Bitcoin. *Journal of Asset Management*, **16**, 365–373.
- CHAN, S., CHU, J., NADARAJAH, S., AND OSTERRIEDER, J. (2017). A statistical analysis of cryptocurrencies. *Journal of Risk and Financial Management*, **10**, 12.
- COECKELBERGH, M. AND REIJERS, W. (2016). Cryptocurrencies as narrative technologies. *ACM SIGCAS Computers and Society*, **45**, 172–178.
- CONLON, T., CORBET, S., AND MCGEE, R. J. (2020). Are cryptocurrencies a safe haven for equity markets? An international perspective from the COVID-19 pandemic. *Research in International Business and Finance*, **54**, 101248.
- CONT, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, **1**, 223.
- DYHRBERG, A. H. (2016a). Bitcoin, gold and the Dollar – A GARCH volatility analysis. *Finance Research Letters*, **16**, 85–92.
- DYHRBERG, A. H. (2016b). Hedging capabilities of Bitcoin. Is it the virtual gold? *Finance Research Letters*, **16**, 139–144.
- ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.
- ENGLE, R. F. AND KRONER, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, **11**, 122–150.
- FAKHFEKH, M. AND JERIBI, A. (2020). Volatility dynamics of crypto-currencies' returns: Evidence from asymmetric and long memory GARCH models. *Research in International Business and Finance*, **51**, 101075.
- GKILLAS, K. AND KATSIAMPA, P. (2018). An application of extreme value theory to cryptocurrencies.

- Economics Letters*, **164**, 109–111.
- GRONWALD, M. (2019). Is Bitcoin a commodity? On price jumps, demand shocks, and certainty of supply. *Journal of International Money and Finance*, **97**, 86–92.
- KANG, S. H., YOON, S.-M., BEKIROU, S., AND UDDIN, G. S. (2020). Bitcoin as hedge or safe haven: Evidence from stock, currency, bond and derivatives markets. *Computational Economics*, **56**, 529–545.
- KLEIN, T., THU, H. P., AND WALTHER, T. (2018). Bitcoin is not the new gold – a comparison of volatility, correlation, and portfolio performance. *International Review of Financial Analysis*, **59**, 105–116.
- MAHESHWARI, R. (2023). Why is the crypto market rising today?
URL: <https://www.forbes.com/advisor/in/investing/cryptocurrency/why-is-crypto-going-up/>
- MAROBHE, M. I. (2022). Cryptocurrency as a safe haven for investment portfolios amid COVID-19 panic cases of Bitcoin, Ethereum and Litecoin. *China Finance Review International*, **12**, 51–68.
- MELKI, A. AND NEFZI, N. (2022). Tracking safe haven properties of cryptocurrencies during the COVID-19 pandemic: A smooth transition approach. *Finance Research Letters*, **46**, 102243.
- MENSI, W., SENSOY, A., ASLAN, A., AND KANG, S. H. (2019). High-frequency asymmetric volatility connectedness between Bitcoin and major precious metals markets. *The North American Journal of Economics and Finance*, **50**, 101031.
- O'DWYER, R. (2015). The revolution will (not) be decentralised: Blockchains. *Commons Transition*, **11**.
- PHILLIP, A., CHAN, J., AND PEIRIS, S. (2019). On long memory effects in the volatility measure of cryptocurrencies. *Finance Research Letters*, **28**, 95–100.
- PIETERS, G. AND VIVANCO, S. (2017). Financial regulations and price inconsistencies across Bitcoin markets. *Information Economics and Policy*, **39**, 1–14.
- POLASIK, M., PIOTROWSKA, A. I., WISNIEWSKI, T. P., KOTKOWSKI, R., AND LIGHTFOOT, G. (2015). Price fluctuations and the use of Bitcoin: An empirical inquiry. *International Journal of Electronic Commerce*, **20**, 9–49.
- RAHEEM, I. D. (2021). COVID-19 pandemic and the safe haven property of Bitcoin. *The Quarterly Review of Economics and Finance*, **81**, 370–375.
- RUBBANIY, G., KHALID, A. A., AND SAMITAS, A. (2021). Are cryptos safe-haven assets during COVID-19? Evidence from wavelet coherence analysis. *Emerging Markets Finance and Trade*, **57**, 1741–1756.
- VARDAR, G. AND AYDOGAN, B. (2019). Return and volatility spillovers between Bitcoin and other asset classes in Turkey: Evidence from VAR-BEKK-GARCH approach. *EuroMed Journal of Business*, **14**, 209–220.
- VINCENT, O. AND EVANS, O. (2019). Can cryptocurrency, mobile phones, and internet herald sustainable financial sector development in emerging markets? *Journal of Transnational Management*, **24**, 259–279.
- WANG, G., TANG, Y., XIE, C., AND CHEN, S. (2019). Is Bitcoin a safe haven or a hedging asset? Evidence from China. *Journal of Management Science and Engineering*, **4**, 173–188.

WU, C. Y., PANDEY, V. K., AND DBA, C. (2014). The value of Bitcoin in enhancing the efficiency of an investor's portfolio. *Journal of Financial Planning*, **27**, 44–52.

YERMACK, D. (2015). Is Bitcoin a real currency? An economic appraisal. In *Handbook of Digital Currency*. Elsevier, 31–43.

A comparative study of ridge-based adaptive weights in penalised quantile regression on variable selection and regularisation

Innocent Mudhombo¹ and Edmore Ranganai²

¹Department of Accountancy, Vaal University of Technology, Vanderbijlpark Campus, South Africa

²Department of Statistics, University of South Africa, Florida Campus, South Africa

We compare the performance of two adaptive weights in the presence of collinearity in a quantile regression (QR) framework. The first adaptive weights are based on the ridge regression β parameter, as compared to the ridge penalised quantile regression (QRR) based parameters. The QRR based adaptive weights have the advantage of having different weights at each regression quantile (RQ) level, in contrast to the ridge regression RR based weights which do not depend on quantile levels. These adaptive weights are used to formulate the adaptive penalised QR procedures, namely, the adaptive RR penalised QR ($QR-AR$), adaptive $LASSO$ penalised QR ($QR-ALASSO$), and the adaptive elastic net penalised QR ($QR-AE-NET$). The performance of the adaptive weights is measured in terms of how the respective $QR-AR$, $QR-ALASSO$ and $QR-AE-NET$ procedures perform in variable selection and regularisation. A simulation study is used to compare the adaptive weights based on their variable selection and regularisation performance in the presence of mixed, moderate, and high collinearity. The RR -based adaptive weights outperform the QRR -based adaptive weights in prediction under the $QR-ALASSO$ scenario. In contrast, the QRR -based adaptive weights dominate the RR -based adaptive weights under the $QR-AE-NET$ scenario. Under the $QR-AR$ scenario, the adaptive weights perform equally. The QRR -based adaptive weights dominate the percentage of correctly fitted models under the $QR-ALASSO$ and $QR-AE-NET$ scenarios.

Keywords: Adaptive elastic net, Adaptive $LASSO$, Adaptive ridge, Adaptive weights, Penalised quantile regression.

1. Introduction

Variable selection and regularisation in quantile regression (QR) have been topical in recent years, especially in the presence of collinearity. The adverse effects of collinearity in regression analysis are wrong signs of parameter estimates, erroneous interpretation of parameter estimates, and estimates with disproportionately large variances, amongst others (Hoerl and Kennard, 1970). The phenomenon of collinearity occurs when at least two predictor variables are intercorrelated, resulting in an almost impossible separation of coefficient influences in the regression equation. In the literature, population characteristics, deficiencies in sampling, and over-defined models are major sources of collinearity (see Gunst and Mason, 1980; Montgomery, 2017; Adkins et al., 2015). These collinearity challenges

Corresponding author: Innocent Mudhombo (innocentm@vut.ac.za)

MSC2020 subject classifications: 62G09, 62F35, 62J07

have been mitigated via variable selection and regularisation, with varying degrees of success. In the literature, to circumvent the problem of collinearity, the ridge regression (*RR*) (Hoerl and Kennard, 1970), the *LASSO* regression (Tibshirani, 1996), and their mixture version, namely the elastic net (*E-NET*) (Zou and Hastie, 2005), have been suggested.

The least absolute deviation (*LAD*) procedure (Norouzirad et al., 2018) is a robust procedure that generalises to *QR* at any quantile level of interest. In the literature, like the *LAD* procedure, the *LASSO* procedure is based on the ℓ_1 -norm penalty; hence, it was conveniently modified to the least absolute deviation *LASSO* (*LAD-LASSO*) and weighted least absolute deviation *LASSO* (*WLAD-LASSO*), which have oracle properties when appropriate tuning parameters are chosen (Arslan, 2012). In a similar fashion, *QR-LASSO* and *QR-E-NET* procedures have been suggested as variable selection and regularisation in the *QR* framework (Ranganai and Mudhombo, 2021). Although *LASSO* regression (Tibshirani, 1996) does parameter shrinkage and variable selection, simultaneously, and is appropriate for variable selection and regularisation, it falls short in the presence of collinearity. *LASSO* tends to over-penalise coefficients, especially in the presence of collinearity, where all coefficients in a group of correlated variables are penalised to zero except one. On the contrary, ridge regression (Hoerl and Kennard, 1970) is far less "greedy" as it tends to select all coefficients and result in a complex model. The *E-NET* (see Zou and Hastie, 2005) was proposed in response to the challenges of the *LASSO* and the *RRs*, and is a compromise between the two procedures. The *LASSO* regularisation method, which has an ℓ_1 -norm penalty, is dominated in prediction performance by the ridge procedure (Zou and Hastie, 2005). The *LASSO* and *E-NET* regularisation procedures have been extended to their adaptive scenarios, namely; the adaptive *LASSO* (*ALASSO*) and adaptive elastic net (*AE-NET*), respectively, as solutions to problems posed by collinearity in data sets (see Zou, 2006; Zou and Zhang, 2009). In literature, to circumvent the problem of collinearity, adaptive penalised variable selection and regularisation procedures are suggested, such as adaptive ridge regression (*AR*) (Frommlet and Nuel, 2016), *ALASSO* (Zou, 2006), and the adaptive elastic net (*AE-NET*) (Zou and Zhang, 2009). The *ALASSO* was proposed by (Zou, 2006), and it allows different tuning parameters for different coefficients. The suggested *ALASSO*, uses ridge regression coefficient estimates to form adaptive weights.

The performance of variable selection and regularisation procedures heavily depends on the appropriate selection of the tuning parameters. For these procedures, the true model is identified consistently depending on the appropriate tuning parameter selection (see Fan and Li, 2001; Zou, 2006). In literature, methods such as C_p , the Akaike information criterion (*AIC*), the Bayesian information criterion (*BIC*), cross-validation (*CV*), and bootstrap have been used for variable selection and choosing tuning parameters in regularisation techniques (Hastie et al., 2009). The C_p , the *AIC*, and the *BIC* methods are estimators of in-sample prediction errors. The basis functions are used in the proportional adjustment of the training error in the C_p criterion, and the *AIC* criterion uses a log-likelihood loss function instead. Unlike the *AIC*, the *BIC* gives preference to uncomplicated models in variable selection over complex ones, which are penalised heavily. In contrast, some out-of-sample estimators of prediction errors include the *CV* and bootstrap methods as examples. The method of *CV* is widely used to choose the tuning parameters (λ_{min}) in the literature. In the regularisation and penalisation techniques, some criteria are used with *CV* criteria to select variables. In the *CV* technique, estimates from the training set are compared to the rest of the data (validation set).

The motivations that undergird this study are as follows:

- We carry out a detailed comparative study of the performances of *RR*-based adaptive weights and *QRR*-based adaptive weights under different levels of collinearity at different distribution scenarios, namely:
 - mixed collinearity (three predictor variables are highly correlated and the other two are not);
 - moderate collinearity (all five predictor variables have moderate correlations);
 - high collinearity (all five predictor variables have high or severe correlations, i.e., above 0.80).
- The adaptive weights are based on the *RR* and *QRR* coefficients. The *RR*-based adaptive weight is a global estimate, as suggested in the literature, compared to the *QRR*-based adaptive weights, which are local. The *QRR*-based adaptive weights are different at each quantile level. The adaptive variable selection and regularisation procedures based on these adaptive weights are the *QR-AR*, *QR-ALASSO*, and *QR-AE-NET* procedures.
- We use simulation studies and an example from the literature to carry out a comparative study of adaptive weights using penalised variable selection and regularisation procedures in the *QR* framework, namely; *QR-AR*, *QR-ALASSO*, and *QR-AE-NET*. A better performance by the regularisation procedure translates to a better performance by the adaptive weights.

The rest of the article is organised as follows. Section 2 reviews the adaptive weights for penalised procedures, namely, *RR*-based adaptive weights and *QRR*-based adaptive weights. In Section 2.1, we review the adaptive penalised *QR* variable selection and regularisation techniques, namely, *QR-AR*, *QR-ALASSO*, and *QR-AE-NET*. Simulations are done in Section 3, with simulation results discussed in Sections 3.1 and 3.2 and examples discussed in Section 3.3. We conclude with a discussion in Section 4.

2. Adaptive penalised quantile regression and regularisation procedures

Consider the linear equation given by

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where y_i is the i th entry of the response vector \mathbf{Y} , \mathbf{x}'_i , the i th row vector of the $n \times p$ design matrix \mathbf{X} , $\boldsymbol{\beta}$ is the vector of parameters to be estimated from the data, and $\epsilon_i \sim F$, the i th error term. The *RR* estimator with an ℓ_2 penalty (Hoerl and Kennard, 1970) for the coefficient vector $\boldsymbol{\beta}$ in (1), is given by the minimisation problem

$$\hat{\boldsymbol{\beta}}_{RR} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^n (y_i - \mathbf{x}'_i \boldsymbol{\beta})^2 + n\lambda \sum_{j=1}^p \beta_j^2, \quad j = 1, 2, \dots, p, \quad i = 1, 2, \dots, n, \quad (2)$$

where λ is a positive tuning parameter in the range $0 < \lambda < 1$, the second term is the penalty term, and $\boldsymbol{\beta}$ is a vector of parameters, found using the ridge trace. The *RR* estimator is the most popular

regularisation procedure that deals with collinearity, though its drawbacks are bias and instability, culminating from its dependence on λ (Muniz and Kibria, 2009). The $\beta(\lambda) \rightarrow \beta_{LS}$ as $\lambda \rightarrow 0$, which is an unbiased estimator of β . The best λ value is when the system stabilises with orthogonal characteristics and the issue of incorrect signs of coefficients and the inflated sum of squared errors (*SSE*) is resolved.

Consider the *RR* solution in (1), thus the first adaptive weight (*RRW*) is given by

$$\omega_j = \left(|\hat{\beta}_{RR_j}| + 1/n \right)^{-\gamma}, \quad j = 1, 2, \dots, p, \quad (3)$$

where $\hat{\beta}_{RR_j}$ is the j th *RR* parameter estimate, and $1/n$ is added to avoid dividing by a near zero term, for $\gamma > 0$. Frommlet and Nuel (2016) proposed the adaptive weights $\omega = (|\hat{\beta}_{RR_j}|^\gamma + \delta^\gamma)^{(\theta-2)/\gamma}$, translating to (3) when $\theta = 1$, $\delta = 1/n$ and $\gamma = 1$.

We introduce the second adaptive weights by first stating a *RR* penalised quantile regression (*QRR*) (Hoerl and Kennard, 1970; Koenker and Bassett Jr, 1978). The *QRR* procedure is the minimisation problem

$$\hat{\beta}(\tau)_{QRR} = \underset{\beta \in R^p}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau |y_i - \mathbf{x}'_i \beta(\tau)| + n\lambda \sum_{j=1}^p \beta_j^2, \quad (4)$$

where $\hat{\beta}(\tau)_{QRR_j}$ is the j th coefficient estimate at the τ th regression quantile (*RQ*) level, and λ is the tuning parameter. The check function,

$$\rho_\tau(u) = \begin{cases} \tau u & \text{if } u \geq 0, \\ (\tau - 1)u & \text{if } u < 0, \end{cases}$$

$i = 1, 2, \dots, n$, denote the re-weighting function of residuals u for $\tau \in (0, 1)$.

The *QRR* coefficients are then used in formulating the *QRR*-based adaptive weight (*QRRW*) (see Mudhombo and Ranganai, 2022) given by

$$\tilde{\omega}_j = \left(|\hat{\beta}(\tau)_{QRR_j}| + 1/n \right)^{-1}, \quad j = 1, 2, \dots, p, \quad (5)$$

where $\tilde{\omega}_j$ are the *QRR*-based adaptive weights at a specified τ quantile level and other terms are defined in (4). The adaptive weights, $\tilde{\omega}_j$ can be adjusted to a particular distribution and to all τ quantile levels.

A penalty in regularisation procedures takes the form of a bridge penalty term: $\sum_{j=1}^p |\hat{\beta}_j|^q$. When $q = 1$ and $q = 2$, the bridge penalty becomes the *LASSO* and *RR* penalties, respectively, as special cases. A combination of the *LASSO* and *RR* penalties results in the *E-NET* penalty, which inherits their respective properties. The inclusion of adaptive weights from (3) and (5) results in the adaptive bridge penalty, $\sum_{j=1}^p \varphi |\hat{\beta}_j|^q$. The *ALASSO* and *AR* penalties are special cases when $q = 1$ and $q = 2$. The combination results in *AE-NET*, where $\varphi \in (\omega_j; \tilde{\omega}_j)$ is the adaptive weight. These adaptive weights can be applied to both the least squares (*LS*) and *QR* scenarios.

2.1 Adaptive penalised quantile regression, regularisation and variable selection

In this section, we summarise adaptive penalised *QR* regularisation procedures. These adaptive penalised *QR* procedures are used to compare the performance of the adaptive weights ω_j and $\tilde{\omega}_j$

(the *LS* and *QR*-based adaptive weights). For further reading on adaptive weights, the reader is referred to *ALASSO* (Zou, 2006) and *AE-NET* (Zou and Zhang, 2009).

We present the adaptive penalised *QR* regularisation and variable selection procedures with adaptive weights presented in (3) and (5). Consider a *QR* with an *AE-NET* penalty denoted by *QR-AE-NET* (see also Zou and Zhang, 2009, for the *LS* version of the *AE-NET* regularisation procedures). The *QR-AE-NET* procedure has both the ℓ_1 and ridge penalties, hence it is an extension of both the adaptive *LASSO* penalised *QR* (*QR-ALASSO*) and *AR* penalised *QR* (*QR-AR*) procedures. The *QR-AE-NET* inherits some attractive properties from both the *QR-ALASSO* and *QR-AR* procedures. The *QR-AE-NET* procedure is given by the minimisation problem

$$\hat{\beta}(\tau) = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^n \rho_{\tau}|y_i - \mathbf{x}'_i \beta(\tau)| + \alpha \lambda \sum_{j=1}^p \varphi |\beta_j| + (1 - \alpha) \lambda \sum_{j=1}^p \varphi \beta_j^2, \quad (6)$$

where φ is one of the adaptive weights ω_j and $\tilde{\omega}_j$, $\alpha \in [0 : 1]$ is mixing parameter resulting in *QR-AR* ($\alpha = 0$) and *QR-ALASSO* ($\alpha = 1$), and λ is the tuning parameter for the two adaptive penalties. In this article, the ℓ_1 and ridge penalties have equal weighting in *QR-AE-NET*, achieved by invoking the mixing parameter $\alpha = 0.50$. The tuning parameter, $\lambda_j = \lambda \varphi$, is varying for $j = 1, 2, \dots, p$ and shrinks coefficients to zero differently. Equations (1)–(5) define the other terms. The *QR-ALASSO* and *QR-AE-NET* procedures inherit the desired optimal minimax bound from *ALASSO* (see Zou, 2006) and the procedures are also robust in the presence of collinearity. Under suitable conditions, the variable selection and regularisation techniques satisfy the sparsity condition, and the distribution converges in limit to a normal distribution in the *QR* scenarios.

3. Simulation study

In this section, we compare the performances of *RR* and *QRR*-based adaptive weights (ω_j and $\tilde{\omega}_j$) under penalised *QR* procedures. These adaptive weights are compared in terms of their ability to improve the performance of *AR*, *ALASSO*, and *AE-NET* penalised *QR* procedures (*QR-AR*, *QR-ALASSO*, and *QR-AE-NET*) in variable selection and regularisation at $\tau \in (0.25, 0.50, 0.75)$ *RQ* levels. The simulation results are summarised in terms of the *MAD* of test errors, the percentage of correctly fitted models, and the average of correct zero coefficients.

3.1 Design scenarios

We consider three design scenarios, namely, the mixed, moderate, and high collinearity design matrices. These simulation design scenarios are simulated as follows:

- (1) Generate the matrix \mathbf{Z} with five variables (Gibbons, 1981), where

$$Z_{ij} \sim N(0, 1) \begin{cases} i = 1, 2, 3, \dots, n, \\ j = 1, 2, 3, 4, 5. \end{cases} \quad (7)$$

- (2) We then generate two design matrices as follows:

$$X_{1ij} = (1 - \theta^2)^{1/2} Z_{ij} + \theta Z_{i5} \begin{cases} i = 1, 2, 3, \dots, n, \\ j = 1, 2, 3, \end{cases} \quad (8)$$

and

$$X_{2ij} = (1 - \theta^{*2})^{1/2} Z_{ij} + \theta^* Z_{i5} \begin{cases} i = 1, 2, 3, \dots, n, \\ j = 4, 5. \end{cases} \quad (9)$$

(3) Form the 60×5 design matrix $\mathbf{X} = (\mathbf{X}_{1ij}, \mathbf{X}_{2ij})$, resulting in severe/high collinearity ($\theta = \theta^* = 0.90$), moderate collinearity ($\theta = \theta^* = 0.7$), and mixed collinearity ($\theta = 0.90$ and $\theta^* = 0.1$), where θ is the theoretical correlation between any pair of the first three variables, and θ^* is theoretical correlation between X_4 and X_5 . The coefficients are such that $\beta_0 = 0$, and β is the eigenvector corresponding to the largest eigenvalue of $\mathbf{X}^{*'} \mathbf{X}^*$, where \mathbf{X}^* is a standardised design matrix and $\mathbf{X}^{*'} \mathbf{X}^*$ is in correlation form.

(4) Lastly, we generate the response variable by

$$y_i = \mathbf{x}'_i \beta + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (10)$$

where $n = 60$, $\epsilon_i \sim t_d$ is the error term (d is the degrees of freedom, where $d \in (6; 20)$), and \mathbf{x}'_i is the i th row of the design matrix \mathbf{X} . The coefficient vector β is given by $\beta = (0.9, 0, 0, 0, 0.5)$ for the mixed collinearity scenario, $\beta = (0.9, 0, 0.7, 0, 0.6)$ for the moderate collinearity scenario, and $\beta = (0.9, 0.7, 0, 0, 0.6)$ for the high collinearity scenario. QR is robust to outliers since RQ s influence functions are bounded in the response variable and QR is designed to handle heavy-tailed distributions, such as t_d . We employed 200 simulation runs and 10-fold cross-validation to obtain the tuning parameters.

We use the `hqreg` R package (<http://cloud.r-project.org/package=hqreg>) for our simulations and data analysis (Yi, 2017). The `hqreg` program chooses the optimal λ (minimum λ) by the K -fold CV criterion (see also Ranganai and Mudhombo, 2021; Mudhombo and Ranganai, 2022).

3.2 Results

We compare the performance of two adaptive weights (ω_j and $\tilde{\omega}_j$) applied to penalised adaptive QR techniques in variable selection and regularisation in the presence of collinearity. The simulated results are summarised and discussed in this section. Tables 1, 2, 3 and 4 show the performance of two adaptive weights (ω_j and $\tilde{\omega}_j$) when applied to different variable selection and regularisation procedures ($QR-AR$, $QR-ALASSO$, and $QR-AE-NET$) using MAD of test errors, percentage of correctly fitting the models, and the average of correct zero coefficients at $\tau \in (0.25, 0.50, 0.75)$ RQ levels and $d = (6; 20)$ degrees of freedom (see also Figure 1). The performance of these penalised QR techniques gauges the performance of corresponding adaptive weights. The MAD of test errors is given by $MAD = 1.4826 (\text{median} | e_i - \text{median} \{e_i\} |)$, for $1 \leq i \leq n$.

Mixed collinearity scenario

The performance of the adaptive weights under a mixed collinearity design matrix scenario is shown in Tables 1, 2, 3, 4, and Figure 1. Under $QR-ALASSO$ and $QR-AE-NET$, the adaptive weights perform equally across all RQ levels 50% of the time. However, ω_j outperforms $\tilde{\omega}_j$ in prediction 33% of the time under the $QR-ALASSO$ technique and vice versa under the $QR-AE-NET$ procedure. When applied to $QR-ALASSO$, $\tilde{\omega}_j$ outperform ω_j in correctly fitting the model (83% of the time), and conversely, ω_j outperform $\tilde{\omega}_j$ 50% of the time under the $QR-AE-NET$ procedure. However,

the two adaptive weights perform equally well in correctly fitting the models 100% of the time under the *QR-AR* technique.

Moderate collinearity scenario

Tables 1, 2, 3, 4, and Figure 1 show the performance of adaptive weights ω_j and $\tilde{\omega}_j$ in terms of prediction and percentage of correctly fitting the models when moderate collinearity is present in the data. The adaptive weights $\tilde{\omega}_j$ dominate the prediction performance 50% of the time when applied to the *QR-ALASSO* technique. Under the *QR-AE-NET* and *QR-AR* procedures, the adaptive weights have similar predictive performance for 50% and 67% of the time, respectively, for $d \in (6; 20)$ degrees of freedom at $\tau \in (0.25, 0.50, 0.75)$ *RQ* levels. When applied to *QR-AE-NET* and *QR-AR* techniques, the adaptive weights $\tilde{\omega}_j$ outperform ω_j 67% and 50% of the time in correctly fitting the models, respectively (50% for ω_j under *QR-AE-NET* scenario). However, under the *QR-AR* technique, there is no difference in predictive performance or percentage of correctly fitted models of the adaptive weights 67% and 100% of the time, respectively.

High collinearity scenario

Under *QR-ALASSO* and *QR-AR* scenarios, the predictive performance of the adaptive weights is the same 67% and 67% of the time, respectively ($\tilde{\omega}_j$ dominate 33% of the time). The exception is under the *QR-AE-NET* procedure, where $\tilde{\omega}_j$ dominate 67% of the time in predictive performance (see Tables 1, 2, 3, and 4). In the percentage of correctly fitted models, $\tilde{\omega}_j$ dominate 83% and 50% of the time under the *QR-ALASSO* and *QR-AE-NET* scenarios, respectively. The two adaptive weights have similar performance in terms of the percentage of correctly fitted models 100% of the time under the *QR-AR* technique.

Remark 1. In Tables 1, 2, 3, and 4, the performance of the adaptive weights ω_j and $\tilde{\omega}_j$ in penalised *QR* procedures at $\tau = (0.25, 0.50)$ *RQ* levels. A better performance by the penalised *QR* procedures indicates a better performance by the corresponding adaptive weights.

3.3 Examples

Under three adaptive *QR* procedures for variable selection and regularisation, we compare the performance of adaptive weights ω_j and $\tilde{\omega}_j$ namely, *QR-ALASSO* and *QR-AE-NET*, using the Jet-Turbine Engine (Montgomery et al., 2009) data set. The 40 observation Jet-Turbine Engine data are known to be correlated (see also Bagheri and Midi, 2012). In this data set, primary speed of rotation (X_1), secondary speed of rotation (X_2), fuel flow rate (X_3), pressure (X_4), exhaust temperature (X_5), and ambient temperature at time of test (X_6) are predictor variables with a response variable (Y). We generate the response variable by $Y_i = \mathbf{X}'_i \boldsymbol{\beta} + \epsilon_i$, where $\epsilon_i \sim t_d$ ($d \in (6; 20)$) is the error term, \mathbf{X}'_i is the i th row of the design matrix \mathbf{X} which is in correlation form, and $\boldsymbol{\beta} = (0, 0, 0, 6, 0, -3)'$ is the vector of parameters. Results are reported only at $\tau \in 0.25, 0.50$ since similar results were found when $\tau = 0.75$ *RQ* level.

The results of the estimated *QR* β s of *QR-ALASSO* and *QR-AE-NET* procedures based on adaptive weights ω_j and $\tilde{\omega}_j$, and coefficient biases are presented in Tables 5 and 6. Zero coefficients are shrunk to zero/near zero in both scenarios (100% of the time) for all adaptive weights. The adaptive weights ω_j yields marginally better results than $\tilde{\omega}_j$ under *QR-ALASSO* when $d \in (6, 20)$ at $\tau = 0.25$ *RQ* level. At the same *RQ* level, $\tilde{\omega}_j$ yields marginally better results than ω_j under the

Table 1. Performance of adaptive weights in *QR-ALASSO* at mixed, moderate, and high collinearity scenarios under the heavy-tailed t-distributions when $d = 6$ & $d = 20$ degrees of freedom.

	Adaptive weight	Median (MAD) test error	Correctly fitted	Average no. of		Median(λ)
				correct zero	incorrect zero	
$d = 6, \tau = 0.25$						
Mixed collinearity	ω_j	0.78(1.17)	71.00	2.94	0.29	0.01
	$\tilde{\omega}_j$	0.76(1.18)	58.00	2.90	0.42	0.00
Moderate collinearity	ω_j	0.82(1.27)	58.50	1.70	0.20	0.01
	$\tilde{\omega}_j$	0.81(1.29)	38.50	1.44	0.27	0.01
High collinearity	ω_j	0.83(1.30)	51.00	1.94	0.50	0.01
	$\tilde{\omega}_j$	0.83(1.30)	40.00	1.94	0.66	0.01
$d = 6, \tau = 0.50$						
Mixed collinearity	ω_j	-0.04(1.16)	74.00	2.98	0.27	0.01
	$\tilde{\omega}_j$	-0.04(1.16)	77.50	2.96	0.22	0.01
Moderate collinearity	ω_j	0.01(1.22)	58.50	1.66	0.13	0.01
	$\tilde{\omega}_j$	0.00(1.22)	79.00	1.87	0.09	0.01
High collinearity	ω_j	0.02(1.27)	52.50	1.94	0.43	0.01
	$\tilde{\omega}_j$	0.02(1.27)	58.00	1.98	0.41	0.02
$d = 6, \tau = 0.75$						
Mixed collinearity	ω_j	-0.88(1.24)	58.00	2.96	0.52	0.01
	$\tilde{\omega}_j$	-0.88(1.24)	60.00	2.95	0.48	0.01
Moderate collinearity	ω_j	-0.81(1.24)	54.00	1.68	0.25	0.01
	$\tilde{\omega}_j$	-0.80(1.26)	57.00	1.68	0.20	0.01
High collinearity	ω_j	-0.78(1.31)	48.00	1.87	0.45	0.01
	$\tilde{\omega}_j$	-0.77(1.30)	52.50	1.83	0.39	0.01
$d = 20, \tau = 0.25$						
Mixed collinearity	ω_j	0.75(1.13)	62.00	2.98	0.39	0.01
	$\tilde{\omega}_j$	0.73(1.14)	65.00	2.92	0.30	0.01
Moderate collinearity	ω_j	0.71(1.23)	47.50	1.42	0.08	0.01
	$\tilde{\omega}_j$	0.72(1.23)	47.50	1.42	0.06	0.01
High collinearity	ω_j	0.25(1.16)	61.00	1.85	0.29	0.01
	$\tilde{\omega}_j$	0.74(1.16)	66.50	1.93	0.30	0.01
$d = 20, \tau = 0.50$						
Mixed collinearity	ω_j	0.00(1.14)	69.00	2.97	0.30	0.01
	$\tilde{\omega}_j$	0.00(1.13)	73.00	2.98	0.27	0.01
Moderate collinearity	ω_j	-0.02(1.18)	53.00	1.51	0.03	0.01
	$\tilde{\omega}_j$	-0.01(1.19)	65.00	1.66	0.04	0.01
High collinearity	ω_j	0.00(1.18)	70.50	1.88	0.21	0.01
	$\tilde{\omega}_j$	0.00(1.17)	76.50	1.91	0.16	0.02

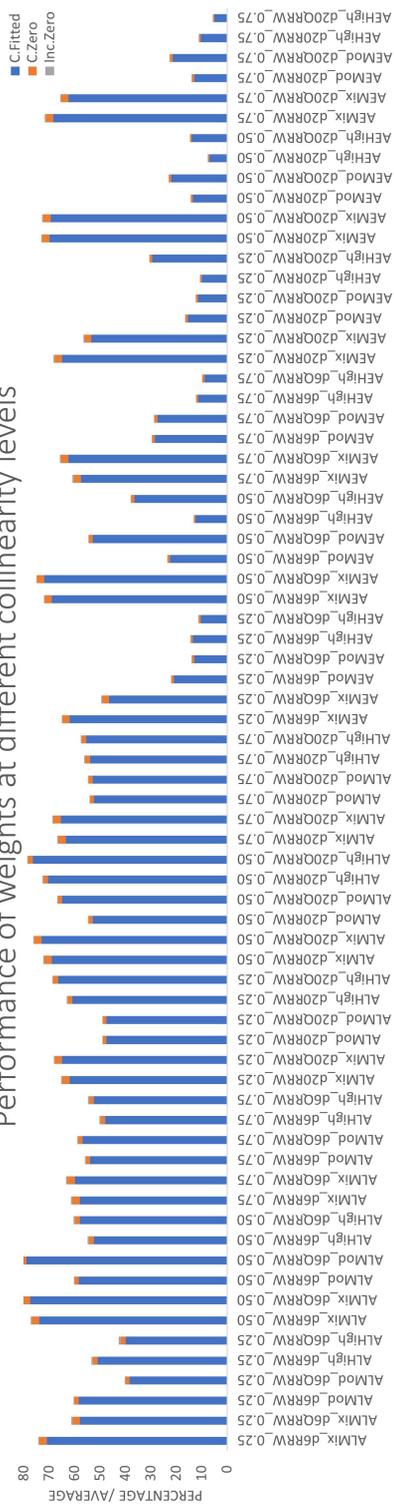
Table 2. Performance of adaptive weights in *QR-AE-NET* at mixed, moderate, and high collinearity scenarios under the heavy-tailed t-distributions when $d = 6$ & $d = 20$ degrees of freedom.

	Adaptive weight	Median (MAD) test error	Correctly fitted	Average no. of		Median(λ)
				correct zero	incorrect zero	
Mixed collinearity	ω_j	0.81(1.21)	62.00	2.80	0.26	0.01
	$\tilde{\omega}_j$	0.80(1.21)	46.50	2.70	0.36	0.01
Moderate collinearity	ω_j	0.81(1.31)	21.00	0.94	0.04	0.02
	$\tilde{\omega}_j$	0.82(1.31)	13.00	0.86	0.05	0.02
High collinearity	ω_j	0.85(1.31)	13.50	0.74	0.01	0.02
	$\tilde{\omega}_j$	0.85(1.30)	10.50	0.63	0.01	0.02
$d = 6, \tau = 0.50$						
Mixed collinearity	ω_j	-0.03(1.19)	69.00	2.84	0.20	0.01
	$\tilde{\omega}_j$	-0.03(1.18)	72.00	2.88	0.20	0.01
Moderate collinearity	ω_j	0.00(1.25)	22.50	0.97	0.01	0.02
	$\tilde{\omega}_j$	-0.01(1.25)	53.00	1.46	0.01	0.02
High collinearity	ω_j	0.01(1.28)	12.50	0.68	0.00	0.02
	$\tilde{\omega}_j$	0.02(1.27)	36.50	1.33	0.00	0.03
$d = 6, \tau = 0.75$						
Mixed collinearity	ω_j	-0.90(1.26)	57.50	2.86	0.44	0.01
	$\tilde{\omega}_j$	-0.89(1.25)	62.50	2.90	0.42	0.01
Moderate collinearity	ω_j	-0.83(1.29)	28.50	1.04	0.03	0.02
	$\tilde{\omega}_j$	-0.81(1.28)	28.50	1.17	0.02	0.02
High collinearity	ω_j	-0.79(1.33)	11.50	0.70	0.00	0.02
	$\tilde{\omega}_j$	-0.77(1.32)	9.00	0.80	0.00	0.02
$d = 20, \tau = 0.25$						
Mixed collinearity	ω_j	0.75(1.15)	65.00	2.92	0.31	0.01
	$\tilde{\omega}_j$	0.74(1.16)	53.50	2.69	0.25	0.01
Moderate collinearity	ω_j	0.71(1.24)	15.50	0.86	0.01	0.02
	$\tilde{\omega}_j$	0.71(1.25)	11.50	0.70	0.01	0.02
High collinearity	ω_j	0.75(1.15)	10.00	0.60	0.01	0.02
	$\tilde{\omega}_j$	0.75(1.15)	29.50	1.05	0.00	0.02
$d = 20, \tau = 0.50$						
Mixed collinearity	ω_j	0.01(1.14)	70.00	2.90	0.23	0.01
	$\tilde{\omega}_j$	-0.01(1.14)	70.00	2.95	0.26	0.01
Moderate collinearity	ω_j	-0.01(1.18)	13.50	0.80	0.00	0.02
	$\tilde{\omega}_j$	-0.02(1.18)	22.00	0.95	0.00	0.02
High collinearity	ω_j	0.00(1.17)	7.00	0.53	0.00	0.02
	$\tilde{\omega}_j$	0.00(1.16)	14.00	0.67	0.00	0.03

Table 3. Performance of adaptive weights of the QR - AR procedure at mixed, moderate, and high collinearity scenarios under the heavy-tailed t-distributions when $d = 6$ & $d = 20$ degrees of freedom.

	Adaptive weight	Median (MAD) test error	Correctly fitted	Average no. of		Median(λ)
				correct zero	incorrect zero	
$d = 6, \tau = 0.25$						
Mixed collinearity	ω_j	0.80(1.26)	0.50	0.56	0.00	0.02
	$\tilde{\omega}_j$	0.80(1.25)	0.50	0.58	0.01	0.01
Moderate collinearity	ω_j	0.87(1.40)	0.00	0.01	0.00	0.05
	$\tilde{\omega}_j$	0.88(1.40)	0.00	0.02	0.00	0.07
High collinearity	ω_j	0.90(1.35)	0.00	0.00	0.00	0.05
	$\tilde{\omega}_j$	0.91(1.37)	0.00	0.00	0.00	0.07
$d = 6, \tau = 0.50$						
Mixed collinearity	ω_j	-0.04(1.24)	0.00	0.49	0.00	0.03
	$\tilde{\omega}_j$	-0.05(1.24)	0.00	0.48	0.00	0.04
Moderate collinearity	ω_j	-0.01(1.33)	0.00	0.01	0.00	0.07
	$\tilde{\omega}_j$	-0.01(1.33)	0.00	0.01	0.00	0.07
High collinearity	ω_j	0.01(1.31)	0.00	0.00	0.00	0.06
	$\tilde{\omega}_j$	0.02(1.31)	0.00	0.00	0.00	0.08
$d = 6, \tau = 0.75$						
Mixed collinearity	ω_j	-0.91(1.29)	0.50	0.53	0.01	0.03
	$\tilde{\omega}_j$	-0.91(1.29)	1.00	0.55	0.01	0.04
Moderate collinearity	ω_j	-0.91(1.39)	0.00	0.01	0.00	0.06
	$\tilde{\omega}_j$	-0.91(1.39)	0.00	0.01	0.00	0.05
High collinearity	ω_j	-0.84(1.39)	0.00	0.00	0.00	0.05
	$\tilde{\omega}_j$	-0.83(1.39)	0.00	0.00	0.00	0.05
$d = 20, \tau = 0.25$						
Mixed collinearity	ω_j	0.78(1.20)	0.50	0.64	0.00	0.03
	$\tilde{\omega}_j$	0.79(1.20)	0.50	0.62	0.01	0.03
Moderate collinearity	ω_j	0.78(1.29)	0.00	0.04	0.00	0.05
	$\tilde{\omega}_j$	0.79(1.27)	0.00	0.04	0.00	0.05
High collinearity	ω_j	0.80(1.20)	0.00	0.00	0.00	0.06
	$\tilde{\omega}_j$	0.80(1.20)	0.00	0.00	0.00	0.07
$d = 20, \tau = 0.50$						
Mixed collinearity	ω_j	0.00(1.18)	0.00	0.56	0.00	0.03
	$\tilde{\omega}_j$	0.00(1.18)	0.0	0.55	0.00	0.04
Moderate collinearity	ω_j	-0.02(1.25)	0.00	0.02	0.00	0.07
	$\tilde{\omega}_j$	-0.02(1.24)	0.00	0.01	0.00	0.08
High collinearity	ω_j	0.00(1.20)	0.00	0.00	0.00	0.07
	$\tilde{\omega}_j$	0.00(1.19)	0.00	0.00	0.00	0.08

Performance of weights at different collinearity levels



Comparison of weights using MAD of test errors

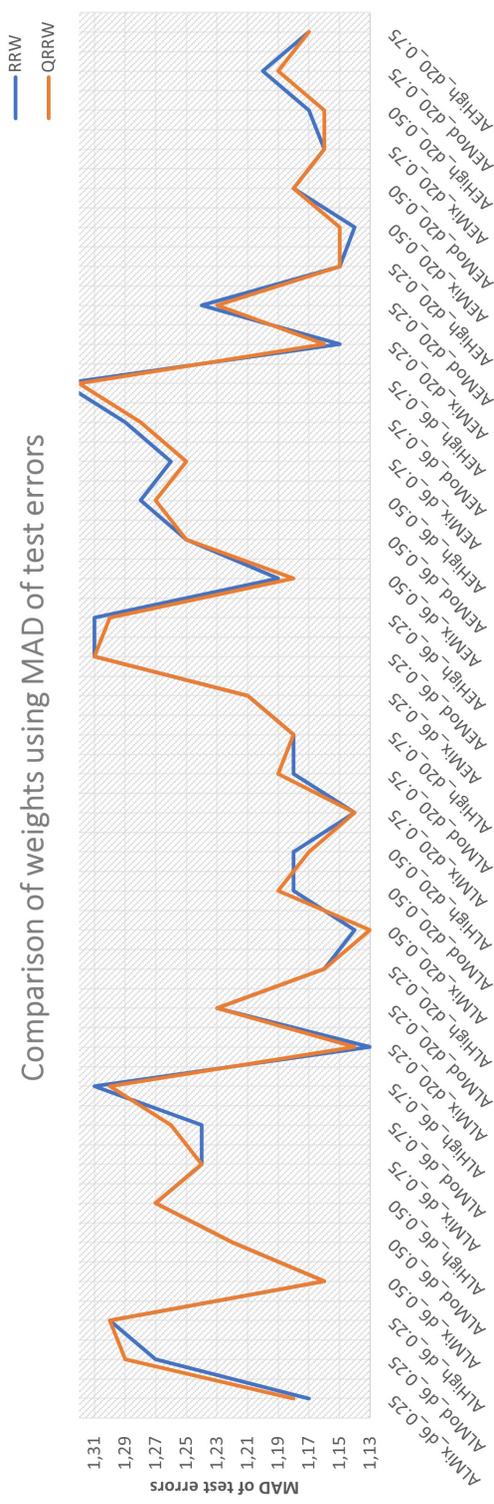


Figure 1. The stacked bar chart shows the performance of the weights ω_j and $\tilde{\omega}_j$ at different collinearity levels. For each pair of stacked bar charts, the first stacked bar represents the performance of ω_j (RRW), and the second represents the performance of $\tilde{\omega}_j$ (QRRW). The second graph shows the performance of the two weights, where the blue line graph is for ω_j and the red one is for $\tilde{\omega}_j$.

Table 4. Performance of adaptive weights of *QR-ALASSO*, *QR-AE-NET*, and *QR-AR* procedures at mixed, moderate, and high collinearity scenarios under the heavy-tailed *t*-distributions when $d = 20$ degrees of freedom.

	Adaptive weight	Median (MAD) test error	Correctly fitted	Average no. of		Median(l)
				correct zero	incorrect zero	
<i>QR-ALASSO</i> , $d = 20$, $\tau = 0.75$						
Mixed collinearity	ω_j	-0.74(1.14)	63.50	2.96	0.38	0.01
	$\tilde{\omega}_j$	-0.73(1.14)	65.50	2.92	0.32	0.00
Moderate collinearity	ω_j	-0.78(1.19)	52.50	1.49	0.08	0.01
	$\tilde{\omega}_j$	-0.79(1.19)	53.00	1.58	0.13	0.01
High collinearity	ω_j	-0.74(1.18)	54.00	1.86	0.37	0.01
	$\tilde{\omega}_j$	-0.74(1.18)	55.50	1.80	0.31	0.01
<i>QR-AE-NET</i> , $d = 20$, $\tau = 0.75$						
Mixed collinearity	ω_j	-0.75(1.16)	68.50	2.92	0.29	0.01
	$\tilde{\omega}_j$	-0.75(1.16)	62.50	2.81	0.26	0.00
Moderate collinearity	ω_j	-0.78(1.20)	13.00	0.83	0.01	0.02
	$\tilde{\omega}_j$	-0.78(1.19)	21.50	1.00	0.02	0.02
High collinearity	ω_j	-0.75(1.17)	10.50	0.57	0.01	0.02
	$\tilde{\omega}_j$	-0.75(1.17)	5.00	0.44	0.01	0.02
<i>QR-ARIDGE</i> , $d = 20$, $\tau = 0.75$						
Mixed collinearity	ω_j	-0.76(1.19)	0.00	0.51	0.00	0.03
	$\tilde{\omega}_j$	-0.75(1.19)	0.00	0.55	0.00	0.01
Moderate collinearity	ω_j	-0.85(1.24)	0.00	0.03	0.00	0.05
	$\tilde{\omega}_j$	-0.85(1.24)	0.00	0.02	0.00	0.07
High collinearity	ω_j	-0.80(1.21)	0.00	0.00	0.00	0.05
	$\tilde{\omega}_j$	-0.80(1.21)	0.00	0.00	0.00	0.04

Table 5. Estimated coefficients and biases for the Jet-Turbine Engine data set with $d = 6$.

Adaptive weight	β	$\tau = 0.25$		$\tau = 0.50$	
		QR-ALASSO $\beta(\text{Bias})$	QR-AE-NET $\beta(\text{Bias})$	QR-ALASSO $\beta(\text{Bias})$	QR-AE-NET $\beta(\text{Bias})$
ω_j	-0.72	-3.20(2.48)	-11.08(10.36)	35.71(-36.43)	35.71(-36.43)
	0.00	0.00(0.00)	0.04(-0.04)	0.01(-0.01)	0.01(0.01)
	0.00	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
	0.00	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
	6.00	6.00(0.00)	5.62(0.38)	5.95(0.05)	5.95(0.05)
	0.00	0.00(0.00)	-0.03(0.03)	0.01(-0.01)	0.01(-0.01)
	-3.00	-2.97(-0.03)	-3.11(0.11)	-2.95(-0.05)	-2.95(-0.05)
$\tilde{\omega}_j$	0.00	-7.30(7.30)	-11.51(11.51)	35.71(-35.71)	35.71(-35.71)
	0.00	0.00(0.00)	0.04(-0.04)	0.01(-0.01)	0.01(-0.01)
	0.00	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
	0.00	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
	6.00	6.00(0.00)	5.62(0.38)	5.96(0.04)	5.95(0.05)
	0.00	0.00(0.00)	-0.03(0.03)	0.01(-0.01)	0.01(-0.01)
	-3.00	-2.93(0.07)	-3.10(0.10)	-2.95(-0.05)	-2.95(-0.05)

¹ The coefficients are estimated at $\tau = (0.25, 0.50)$ RQ levels for each of the adaptive weights ω_j and $\tilde{\omega}_j$ of the penalised QR procedures.

QR - AE - NET scenario. However, at $\tau = 0.50$ and $d \in (6, 20)$, the two adaptive weights perform the same.

4. Discussion

This article compared the QRR -based adaptive weights $\tilde{\omega}_j$ and the RR -based adaptive weights ω_j . These adaptive weights are used to formulate some variable selection and regularisation procedures (QR - $ALASSO$, QR - AE - NET , and QR - AR). The adaptive weights $\tilde{\omega}_j$ have the advantage that each weight is different at each RQ level as compared to constant weights for all quantile levels in the case of ω_j (Mudhombo and Ranganai, 2022).

A simulation study is used to compare the adaptive weights based on their performance in the mixed, moderate, and high collinearity scenarios. We compare the performance of the adaptive weights ω_j and $\tilde{\omega}_j$ by checking the performance of the QR - AR , QR - $ALASSO$ and QR - AE - NET procedures in variable selection and regularisation.

In the presence of mixed collinearity (a combination of very high and very low collinearity), the adaptive weights $\tilde{\omega}_j$ outperform the weights ω_j at the median quantiles, while the latter is better in the lower quantiles in terms of prediction under QR - $ALASSO$ procedure. The QRR -based adaptive weights are superior in correctly fitting models and in correctly shrinking zero coefficients. When the QR - AE - NET procedure is used, $\tilde{\omega}_j$ outperforms ω_j in prediction. The adaptive weights perform the same in prediction under the QR - AR scenario.

In the moderate collinearity situation under the QR - $ALASSO$ scenario, the two adaptive weights

Table 6. Estimated coefficients and biases for the Jet-Turbine Engine data set with $d = 20$.

Adaptive weight	β	$\tau = 0.25$		$\tau = 0.50$	
		QR-ALASSO $\beta(\text{Bias})$	QR-AE-NET $\beta(\text{Bias})$	QR-ALASSO $\beta(\text{Bias})$	QR-AE-NET $\beta(\text{Bias})$
ω_j	-0.72	1.06(-1.78)	24.17(-24.89)	-35.37(34.65)	-35.37(34.65)
	0.00	0.00(0.00)	0.04(-0.04)	0.02(-0.02)	0.02(-0.02)
	0.00	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
	0.00	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
	6.00	5.93(0.07)	5.55(0.45)	5.84(0.16)	5.84(0.16)
	0.00	0.01(-0.01)	0.01(-0.01)	0.00(0.00)	0.00(0.00)
	-3.00	-3.11(0.11)	-3.22(0.22)	-3.16(0.16)	-3.16(0.16)
$\tilde{\omega}_j$	0.00	-4.03(4.03)	54.84(-54.84)	-35.37(35.37)	-35.37(35.37)
	0.00	0.00(0.00)	0.04(-0.04)	0.02(-0.02)	0.02(-0.02)
	0.00	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
	0.00	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
	6.00	5.92(0.08)	5.56(0.44)	5.84(0.16)	5.84(0.16)
	0.00	0.01(-0.01)	0.01(-0.01)	0.00(0.00)	0.00(0.00)
	-3.00	-3.14(0.14)	-3.23(0.22)	-3.16(0.16)	-3.16(0.16)

¹ The coefficients are estimated at $\tau = (0.25, 0.50)$ RQ levels for each of the adaptive weights ω_j and $\tilde{\omega}_j$ of the penalised QR procedures.

perform similarly in prediction performance. Although ω_j performs better in correctly fitting models and correctly shrinking zero coefficients at lower quantile levels, $\tilde{\omega}_j$ performs better at $\tau = 0.50$. The QR - AE - NET scenario shows the RR -based adaptive weights outperforming the QRR -based adaptive weights in the majority of cases in prediction, though $\tilde{\omega}_j$ is better at correctly fitting models.

The adaptive weights have similar prediction performance most of the time in the presence of high collinearity, although $\tilde{\omega}_j$ is better at correctly fitting models in the QR - $ALASSO$ scenario. The adaptive weights are comparatively similar in the percentage of correctly fitted models in all scenarios.

References

- ADKINS, L. C., WATERS, M. S., HILL, R. C., ET AL. (2015). Collinearity diagnostics in gretl. *Economics Working Paper Series*, **1506**, 1–28.
- ARSLAN, O. (2012). Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression. *Computational Statistics & Data Analysis*, **56**, 1952–1965.
- BAGHERI, A. AND MIDI, H. (2012). On the performance of the measure for diagnosing multiple high leverage collinearity-reducing observations. *Mathematical Problems in Engineering*, **2012**.
- FAN, J. AND LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348–1360.
- FROMMLET, F. AND NUEL, G. (2016). An adaptive ridge procedure for l_0 regularization. *PLoS One*,

11, e0148620.

- GIBBONS, D. G. (1981). A simulation study of some ridge estimators. *Journal of the American statistical Association*, **76**, 131–139.
- GUNST, R. AND MASON, R. (1980). Regression analysis and its applications.
- HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, volume 2. Springer.
- HOERL, A. E. AND KENNARD, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67.
- KOENKER, R. AND BASSETT JR, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50.
- MONTGOMERY, D. C. (2017). *Design and Analysis of Experiments*. John Wiley & sons.
- MONTGOMERY, D. C., RUNGER, G. C., AND HUBELE, N. F. (2009). *Engineering Statistics*. John Wiley & Sons.
- MUDHOMBO, I. AND RANGANAI, E. (2022). Robust variable selection and regularization in quantile regression based on adaptive-LASSO and adaptive E-NET. *Computation*, **10**, 203.
- MUNIZ, G. AND KIBRIA, B. G. (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics – Simulation and Computation*, **38**, 621–630.
- NOROUZIRAD, M., HOSSAIN, S., AND ARASHI, M. (2018). Shrinkage and penalized estimators in weighted least absolute deviations regression models. *Journal Statistical Computation and Simulation*, **88**, 1557–1575.
- RANGANAI, E. AND MUDHOMBO, I. (2021). Variable selection and regularization in quantile regression via minimum covariance determinant based weights. *Entropy*, **23**, 33.
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, 267–288.
- YI, C. (2017). *hqreg: Regularization Paths for Lasso or Elastic-Net Penalized Huber Loss Regression and Quantile Regression*. R package version 1.4.
URL: <https://CRAN.R-project.org/package=hqreg>
- ZOU, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418–1429.
- ZOU, H. AND HASTIE, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, **67**, 301–320.
- ZOU, H. AND ZHANG, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics*, **37**, 1733–1751.

Bandwidth selection in a generic similarity test for spatial data when applied to unmarked spatial point patterns

Jamie-Lee Nel, René Stander and Inger N. Fabris-Rotelli

Department of Statistics, University of Pretoria

The similarity between spatial point pattern data sets is crucial for evaluating the quality and changes in spatial data. A generic similarity test has been developed that is able to handle any type of spatial data. When comparing unmarked point patterns, the generic test starts by calculating the kernel density estimate which requires a bandwidth value. In this research, we test the similarity between unmarked point patterns using this recently proposed generic similarity test. The focus of this work is to assess the effect that the bandwidth choice has on this similarity test. A simulation study is done to evaluate the effect of the different bandwidths on the similarity test. From the simulation study, it is seen that the similarity test could be sensitive towards the choice of the bandwidth depending on the number of points being compared and whether the points are being compared on the same window or not but is robust in general.

Keywords: Bandwidth, Generic similarity test, Point patterns, Spatial similarity

1. Introduction

Spatial data is data that references a specific location and contains information about variables at that location. It can take on different forms, namely geostatistical data, lattice data, or point patterns (Cressie, 2015). Geostatistical data are measurements of spatial data that have been collected at predetermined locations. Lattice data are observations observed on a subset of a spatial domain. Point patterns are the collection of events that take place in a finite number of locations. Point patterns may be marked or unmarked. If attributes are observed at each location, the data is known as a marked point pattern, and if only the location is known the data is an unmarked point pattern.

Spatial data sets are declared similar when the spatial data sets originate from the same stochastic process in terms of their spatial structure (Borrajó et al., 2020). Spatial point patterns and the similarity between them have become of interest to many researchers and some tests have been proposed, namely work done by Andresen (2009); Alba-Fernández et al. (2016). These tests may be used to determine how similar spatial point patterns of interest and the population at risk are, to compare two spatial point patterns of interest, or to compare the similarity between one event measured at different time points. This research will specifically focus on a recently developed similarity test for unmarked spatial point patterns by Kirsten and Fabris-Rotelli (2021).

Corresponding author: Inger N. Fabris-Rotelli (inger.fabris-rotelli@up.ac.za)
MSC2020 subject classifications: 62H11

Andresen (2009) developed a test that evaluates the similarity between two different point patterns using a non-parametric approach, known as the spatial point pattern test. The test results in a local measure, as well as a global measure, of spatial similarity. The local measure of similarity is used to indicate the locations of significantly higher, significantly lower, and insignificant differences in the concentration of a spatial point pattern. The output of the local measure of the test can be mapped which makes it a popular test to use. In order to perform the test proposed an index of similarity is calculated for each spatial unit, e.g. grid cells. The proportion of spatial units with a similar spatial pattern for both sets of data is represented by the *S*-index, which is the global similarity measure. The spatial point pattern test has been used to test the spatial similarity of crime data by Andresen (2009), Andresen and Linning (2012), Andresen and Malleson (2013a,b, 2014), and Linning (2015).

Kirsten and Fabris-Rotelli (2021) proposed a generic spatial similarity test that can handle more than one type of spatial data. This test consists of three significant steps. First, a pixel image representation of both data sets must be obtained. Secondly, the structural similarity index (SSIM index) is calculated for each pixel (Wang et al., 2004). In the third step, a global similarity index is calculated based on Andresen's *S*-index (Andresen, 2009).

Using the generic spatial similarity test to obtain a pixel image representation of unmarked point patterns, kernel density estimation (KDE) is used. Kirsten and Fabris-Rotelli (2021) used Diggle's bandwidth and focused on how the similarity test handles various types of spatial data. In this research, we specifically apply the similarity test to unmarked point patterns with the focus to investigate the effect of different bandwidths on the performance of the test.

Using the individual locations of sample data, kernel density estimation results in a smooth empirical distribution function (Węglarczyk, 2018). Węglarczyk (2018) explores the different symmetric and asymmetric kernels that can be used in one-dimensional non-spatial data, such as Gaussian, Epanechnikov, biweight, triangular, gamma, and rectangular. These kernel functions can be extended to spatial data as well, in other words, bivariate data. The type of kernel chosen is not of too much importance, however, the chosen bandwidth plays a fundamental role in kernel density estimation. The bandwidth of the kernel is known as the standard deviation of the kernel or it can be seen as the smoothing parameter of the kernel (Kirsten and Fabris-Rotelli, 2021). There are various bandwidths that can be used when estimating the KDE for unmarked point patterns such as Diggle's bandwidth (Berman and Diggle, 1989), likelihood cross-validation method (Loader, 2006) and Scott's rule of thumb (Odell-Scott, 1992).

Selecting the most suitable bandwidth is not an easy task. Kuter et al. (2011) studied the effects of different bandwidth choices and kernel density functions using Turkish fire density mapping based on forest fire records at the forest sub-district level. Heidenreich et al. (2013) did a simulation study to find a data-driven optimal bandwidth focusing on small and moderate sample sizes and smooth densities. They found that the choice of bandwidth does, in fact, matter in terms of the quality of the density estimation. It was found that different bandwidths are preferred in different situations. This brings us back to the problem at hand, to assess the effect of different bandwidths on the robustness of the similarity test proposed by Kirsten and Fabris-Rotelli (2021).

Section 2 will discuss the methodology used to perform the similarity test as well as introduce different possible bandwidths to be considered. In Section 3 the method will be tested using the different bandwidths with a simulation study. Section 4 will discuss the results of the simulation study. Section 5 will be the conclusion.

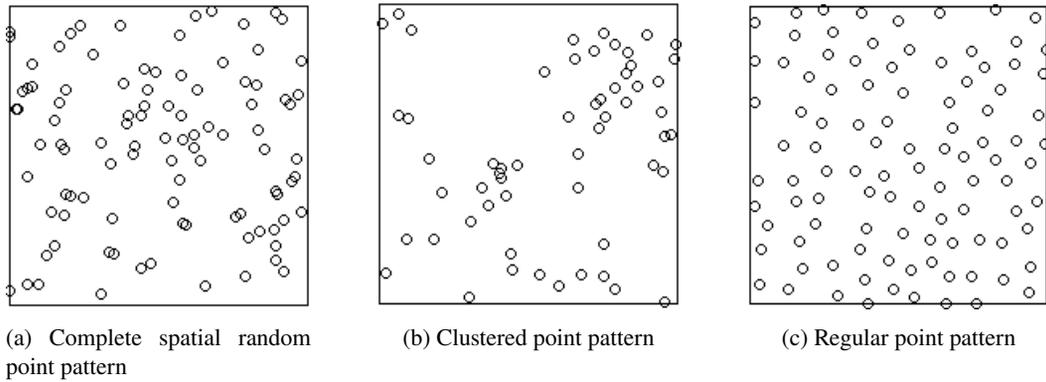


Figure 1. Different types of spatial point patterns.

2. Methodology

2.1 Point pattern theory

A point process $X = \{X_1, X_2, \dots, X_n\}$ with $X_i \in D \subset \mathbb{R}^d$ is a stochastic model governing the location of events in a subset of the spatial domain D (Cressie, 2015). Point processes are stochastic models consisting of irregular point patterns (Illian et al., 2008). A spatial point pattern, $x = \{x_1, x_2, \dots, x_n\}$, is a collection of points giving the observed spatial locations of objects or occurrences (Baddeley et al., 2015). A point pattern is interpreted as a sample from a point process (Illian et al., 2008). In point pattern data analyses $X_i \in D$ would usually be in two or three dimensions. This could be the locations of earthquakes, trees in a forest, road accidents and many more. An example of a point process is a spatial Poisson process (Cox and Isham, 1980).

There are three classifications of point patterns data namely, complete spatial random (CSR), clustered and regular. These are illustrated in Figure 1. A CSR pattern occurs when the locations of the points are randomly distributed in space. A clustered pattern occurs when the points are grouped together in certain regions of space. A regular point pattern occurs when spatial points inhibit each other. If the point pattern is modelled as a Poisson process with parameter λ , where λ is the intensity of the process, then the expected number of points per unit of space for CSR pattern is equal to λ , $E[X] = \lambda$, for a clustered pattern the expected number of points per unit of space is greater than λ , $E[X] > \lambda$, and for a regular pattern it is smaller than λ , $E[X] < \lambda$.

2.2 A spatial similarity test

The following steps outline the generic similarity test proposed by Kirsten and Fabris-Rotelli (2021). For the similarity test two data sets are compared, namely X_1 and X_2 . In the first step a pixel image representation of X_1 and X_2 are created and denoted as Y_1 and Y_2 , which ensures the spatial data types are represented in the same way. The resolution (the number of pixels) of the pixel image should be determined beforehand. A local similarity map is created in the second step, which indicates a local similarity value for each pixel in Y_1 and Y_2 . Finally a similarity percentage from the pixel values in the local similarity map is calculated. The remainder of this section discusses the methods of how the proposed spatial similarity test is applied to unmarked point patterns.

2.2.1 Step 1: Create a pixel image representation

Kernel density estimation is used to obtain the pixel image representation for point patterns. In order to create the pixel image representation, the spatial domain should be divided into an $m \times m$ grid. The centroids of each grid cell are then determined as illustrated in Figure 2. The cells of the grid represents the pixels and the centroids the locations of the centres, u_j .

The spatial locations at the centroids of each of the $M = m^2$ pixels are defined as $u = \{u_1, u_2, \dots, u_M\}$ and $x_i, i = 1, \dots, n$, the point locations of the point pattern. Diggle's corrected density results in a lower mean squared error than similar estimators and is thus used as the density estimate (Baddeley et al., 2015)

$$\tilde{\lambda}^D(u_j) = \sum_{i=1}^n \frac{1}{e(x_i)} \kappa(u_j - x_i), \quad (1)$$

where the kernel is the bivariate Gaussian density $f(d) = (2\pi)^{-1} |\sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2}d\Sigma - 1d'\}$, with $\Sigma = \text{bandwidth} \times I_2$, where I_2 is a 2×2 identity matrix. The Diggle's corrected density includes an edge correction factor, $e(x_i)$. The edge correction factor weighs the points on the boundary less than those within the spatial domain. The edge correction factor in (1) is

$$e(x_i) = \int_D \kappa(x_i - v_k) dv_k, \quad (2)$$

which is estimated using numerical integration. The numerical integration is done by dividing the spatial domain into a finer $g \times g$ grid. The centroids of the $Q = g^2$ grid cells are denoted as the spatial locations $v = \{v_1, v_2, \dots, v_Q\}$. In order to calculate (2) through numerical integration, one must calculate the differences, $d_e = \{d_1, d_2, \dots, d_Q\}$, between the coordinates of each observation in the spatial point pattern, $x_i, i = 1, \dots, n$, and the spatial locations $v_k, k = 1, \dots, Q$, needs to be calculated. The edge correction factor is then calculated as

$$e(x_i) = \frac{\text{area}(D)}{Q} \sum_{k=1}^Q f(d_k), \quad (3)$$

where $f(d_k)$ is the bivariate Gaussian density. An illustration of a point pattern and the resulting pixel image representation for $m = 5$ and $m = 15$ is given in Figure 3.

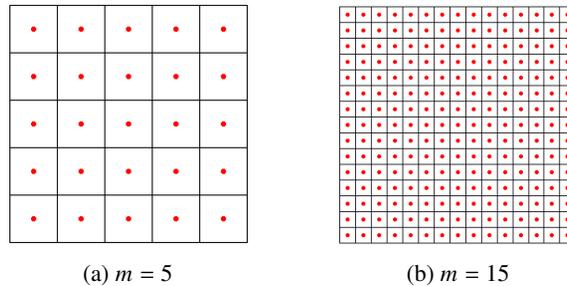


Figure 2. Illustration of how the spatial domain is divided into pixels for two values of parameter m . The u_j are represented by the dots.

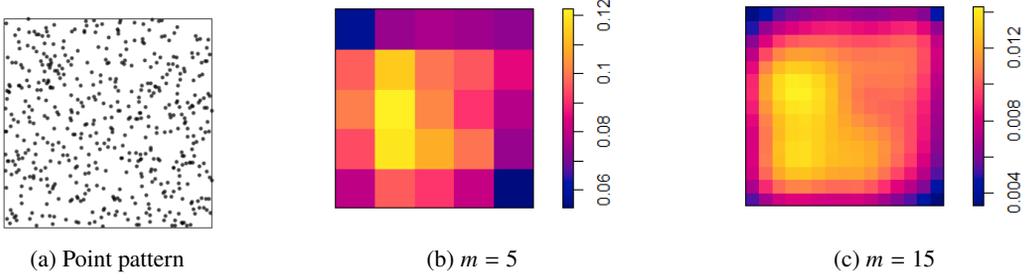


Figure 3. Pixel image representation of an unmarked regular point pattern.

2.2.2 Step 2: Create a local similarity map

Next, a similarity map is created between Y_1 and Y_2 . The SSIM index is used to calculate the local similarity map (Wang et al., 2004). The algorithm uses a sliding window approach in order to move across the image pixel by pixel for both images. The SSIM value is calculated for the centre pixel of each sliding window. In order for the pixel considered to be right at the centre of the sliding window, an odd number of pixels are considered as the length and the width of the sliding window.

The SSIM index (Wang et al., 2004) is calculated as follows,

$$SSIM(y_{1j}, y_{2j}) = [l(y_{1j}, y_{2j})]^\alpha [c(y_{1j}, y_{2j})]^\beta [s(y_{1j}, y_{2j})]^\gamma, \quad (4)$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ and y_{ij} are the values contained in the sliding window j of data set i . Wang et al. (2004) suggests $\alpha = \beta = \gamma = 1$, which ensures equal weight is given to each term. The components of the SSIM value are calculated as follows,

$$\text{Luminance: } l(y_{1j}, y_{2j}) = \frac{2\mu_{y_{1j}}\mu_{y_{2j}} + C_1}{\mu_{y_{1j}}^2 + \mu_{y_{2j}}^2 + C_1},$$

$$\text{Contrast: } c(y_{1j}, y_{2j}) = \frac{2\sigma_{y_{1j}}\sigma_{y_{2j}} + C_2}{\sigma_{y_{1j}}^2 + \sigma_{y_{2j}}^2 + C_2},$$

$$\text{Structure: } s(y_{1j}, y_{2j}) = \frac{\sigma_{y_{1j}, y_{2j}} + C_3}{\sigma_{y_{1j}}\sigma_{y_{2j}} + C_3}.$$

The C_1 , C_2 and C_3 values are the constants which are used to avoid inconsistency (Wang et al., 2004). In literature, the constants are calculated as $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$ and $C_3 = \frac{C_2}{2}$ where $K_1 = 0.01$, $K_2 = 0.03$ and L is the difference between the maximum pixel value and the minimum pixel value from the two images (Wang et al., 2004). The μ_{ij} are the mean values of the Y_i and the σ_{ij} are the standard deviation of the Y_i .

2.2.3 Step 3: Create a global similarity map

Finally the global similarity index is calculated from the local similarity map. The global similarity index is the result of this spatial similarity test. The global similarity index is calculated using the

pixel values in the local similarity map:

$$GS = \frac{1}{M} \sum_{j=1}^M SSIM(u_j), \quad (5)$$

where $SSIM(u_j)$ is the SSIM value for the pixel with centroid u_j and M number of pixels in the pixel image. This provides a mean similarity value instead of a proportion of similar areas as in Andresen's S -index (Andresen, 2009) within the domain which is expected to improve the accuracy.

2.3 Bandwidth selection

Throughout this section x_i will be used as the spatial point pattern locations. It is known that a bandwidth, σ , is the standard deviation or the smoothing parameter of the kernel (Kirsten and Fabris-Rotelli, 2021). The different bandwidths that will be considered to investigate the effect of the bandwidth on the robustness of the similarity test proposed by Kirsten and Fabris-Rotelli (2021) are discussed below.

Diggle's bandwidth

The algorithm used to calculate Diggle's bandwidth uses a method by Berman and Diggle (1989) to compute the quantity

$$M(\sigma) = \frac{MSE(\sigma) - \lambda_d(0)}{\lambda^2}, \quad (6)$$

where σ is the bandwidth, λ the mean intensity, $MSE(\sigma) = E\{[\tilde{\lambda}_\sigma(x_i) - \Lambda(x_i)]^2\}$ is the mean squared error at bandwidth σ and $\tilde{\lambda}_\sigma(x_i) = \frac{N(B_\sigma(x_i))}{|B_\sigma(x_i)|}$. Diggle's bandwidth assumes a stationary Cox process and $\Lambda(x_i)$ is the rate process of the Cox process (Cressie, 2015). $B_\sigma(x_i)$ is the d -dimensional sphere of radius σ centred at x_i and $N(B_\sigma(x_i))$ denotes the number of points of the underlying Cox process in the d -dimensional sphere of radius σ as defined by Berman and Diggle (1989). The bandwidth σ is chosen to minimise the mean square error criterion by direct inspection or numerical integration (Diggle, 1985).

Likelihood cross-validation

This method determines an acceptable bandwidth σ for the kernel density estimate of a point process intensity using a kernel smoothed intensity function for which σ maximises the point process likelihood cross-validation criterion (Loader, 2006)

$$LCV(\sigma) = \sum_{\forall_i} \log(\hat{\lambda}_{-i}(x_i)) - \int_{\forall_D} \hat{\lambda}(u) du,$$

where x_i the point locations of the point pattern, u are the spatial locations of the centroids of the grid, D is the spatial domain and $\hat{\lambda}_{-i}(x_i)$ is leave-one-out kernel-smoothing estimate of the intensity at x_i with smoothing bandwidth σ . The kernel smoothing estimate of the intensity is $\hat{\lambda}(u)$ at a spatial location u with smoothing bandwidth σ (Loader, 2006).

Cronie & van Lieshout's criterion

The bandwidth is selected to reduce the difference between the observation window's area and the sum of the estimated intensities at each point in the point process. Let Ψ be a point process in \mathbb{R}^d , $d \geq 1$, observed inside a non-empty open observation window that is bounded $W \subseteq \mathbb{R}^d$ (Cronie and Van Lieshout, 2018). The likelihood cross-validation criterion is as follows:

$$CvL(\sigma) = \left(|W| - \sum_{y \in \Psi \cap W} \frac{1}{\lambda(x_i)} \right)^2,$$

where $\lambda(x_i)$ is the kernel smoothing estimate of the intensity at x_i , the point locations of the point pattern, with the smoothing bandwidth σ and $|W|$ is the area of W .

Scott's rule of thumb

This bandwidth σ is computed by the rule of thumb of Scott (Odell-Scott, 1992)

$$\sigma \propto n^{\left(\frac{-1}{d+4}\right)},$$

where n is the number of points and d the number of spatial dimensions. In most cases $d = 2$. This rule can be calculated relatively quickly. Compared to Diggle's bandwidth, it often produces a larger bandwidth.

Abramson's adaptive bandwidths for spatial point pattern

The methods used to compute this function were obtained from Abramson (1982) and Hall and Marron (1988). The bandwidth at location u_j is

$$\sigma(u_j) = \frac{\sigma_0}{\tilde{f}(u_j)^{\frac{1}{2}} \gamma},$$

where $\tilde{f}(u_j)$ is a pilot estimate of the spatially varying probability density and the geometric mean of the $\tilde{f}(u_j)^{(-\frac{1}{2})}$ terms evaluated at the data points is $\gamma = \exp\left(\frac{1}{n} \sum_{v_i} \log [f(u_j)^{-\frac{1}{2}}]\right)$. As a result, the global bandwidth σ_0 can be compared to a corresponding fixed bandwidth. The pilot density can either be a pixel image, a fixed bandwidth kernel density estimate using a pilot bandwidth or it can be a different point pattern on the same spatial domain as u_j where the pilot density is then again computed as a fixed-bandwidth kernel density estimate. Abramson's rule is only applied after the pilot density is renormalised to become a probability density for each case.

Bandwidth selection based on window geometry

The bandwidth σ is calculated as a quantile of the distance between two independent random locations in the window. The lower quartile of the distribution is used as the default. Suppose $F(\Sigma)$ is a uniform cumulative distribution function representing the distance between two independent random points in a window, then the value returned is the quantile with probability f . Thus, the value σ such that $F(\sigma) = f$ is the bandwidth (Baddeley et al., 2015).

Stoyan's rule of thumb

Stoyan and Stoyan (1994) proposed a rule of thumb for choosing the smoothing bandwidth. For a general kernel, the smoothing bandwidth is set to $\sigma = c/\sqrt{5\lambda}$, where λ is the estimated intensity of the point pattern and c is a constant. Guan (2007) suggested $c \in (0.1, 0.2)$ with 0.15 as a common choice. Thus, c is chosen as 0.15.

3. Simulation study

A simulation study is conducted in this section to evaluate the performance of the proposed spatial similarity test on different spatial point patterns with various bandwidth choices. The purpose of the simulation study is to generate data, apply the spatial similarity test to the data, and investigate which choice of bandwidth yields the best set of results that are in line with the expected results. The robustness of the choice of the bandwidth value on the similarity test will be evaluated by comparing the similarity scores each bandwidth yields and the expected similarity score.

3.1 Simulation design

We consider spatial point patterns with several characteristics. Hence, various different types of unmarked point patterns and sample sizes are simulated by considering all the combinations of characteristics as outlined in Table 1. The point patterns are also simulated on different windows, various intensities, and sample sizes.

To do this, 230 different spatial data sets are simulated to be used as X_1 . Different techniques are used to obtain the 230 spatial data sets of X_2 which will be compared with X_1 . For each pair of simulated spatial data sets compared, we expect the similarity score equal to the known similarity between the data sets irrespective of the bandwidth used.

The point patterns in the simulation study are simulated such that the point pattern pairs for comparison, X_1 and X_2 , are either 70%, 80%, or 90% identical. To obtain similar spatial point patterns, three different simulation techniques are used. In the first technique the goal is to create noisy patterns. In the second technique, we compare point patterns that have uneven sample sizes. The third simulation technique is only applied to clustered spatial point patterns and seeks to produce spatial point patterns with strong clusters. These are explained in Table 2.

The proposed spatial similarity test by Kirsten and Fabris-Rotelli (2021) is then applied to the 230 different pairs of X_1 and X_2 for each simulation using different bandwidths. The different bandwidths that are considered are those discussed in Section 2.3.

The simulation of the spatial point patterns and the different bandwidths used are built-in R functions from the `spatstat` package (Baddeley et al., 2015). The complete spatial random point patterns are simulated using the `rpoispp` function, the regular spatial point patterns are simulated

Table 1. Summary of parameters considered in the simulation study.

Point pattern	Sample size	Window	Intensity
CSR	Small (± 100 points)	Rectangular	Constant
Regular	Medium (± 500 points)	Polygonal	Non-Constant
Clustered	Large (± 1000 points)		

Table 2. Simulation methods.

	X_1	X_2
Method 1	Simulate CSR, regular, and clustered point patterns.	Replace 10%, 20%, or 30% of data points in X_1 with any other simulated points.
Method 2	Simulate CSR, regular, and clustered point patterns.	Remove 10%, 20%, or 30% of data points from X_1 .
Method 3	Simulate centres as a regular point pattern and simulate clusters as discs around these points.	Replace 10%, 20%, or 30% of data points with simulated data points to be contained within these clusters.

using the `rSSI` function and the clustered spatial point patterns are simulated using the `rMatClust` function (Baddeley et al., 2015). The built-in functions in R for the bandwidths used are, `bw.diggle`, `bw.ppl`, `bw.CvL`, `bw.scott`, `bw.abram.ppp`, `bw.frac` and `bw.stoyan` (Baddeley et al., 2015). Summary statistics of the results obtained from the simulation study are given in Table 3. A constant intensity refers to a homogeneous point pattern and a non-constant intensity refers to an inhomogeneous point pattern.

The bandwidth σ is calculated as a quantile of the distance between two independent random locations in the window. The lower quartile of the distribution is used as the default. Suppose $F(\sigma)$ is a uniform cumulative distribution function representing the distance between two independent random points in a window, then the value returned is the quantile with probability f . Thus, the value σ such that $F(\sigma) = f$ is the bandwidth.

4. Simulation results and discussion

For the first method of simulations, it can be seen that the method performs quite well over all of the different bandwidths except for bandwidth selection based on window geometry. Bandwidth selection based on window geometry yields higher similarity values than expected where the means are 0.8973, 0.9283, and 0.9686, and the medians are 0.9330, 0.9572, and 0.9872. This might be because of the default quartile value, $f = \frac{1}{4}$, used by R. The reason being if $F(\sigma)$ is a uniform cumulative distribution function representing the distance between two independent random points in a window, then the bandwidth is calculated as the distance between two independent random points in the window such that $F(\sigma) = \frac{1}{4}$. The results will be different if the value of f is changed. Stoyan's rule of thumb yields a similarity value slightly lower than expected and its coefficient of variation is the highest. Its variance for 90% identical point patterns is slightly higher than the rest of the bandwidths used but for 70% and 80% identical point patterns the variance is quite similar. This might be because it occasionally yields a negative similarity value. The reason for the negative similarity values might be because the SSIM value is bounded between -1 and 1 . This occurred when CSR point patterns were simulated on a polygonal window with a homogeneous intensity. The similarity values for 70% identical point patterns for Scott's rule of thumb, likelihood cross-validation, Cronie & van Lieshout's criterion and Abramson's adaptive bandwidths were closer to 80% than 70%. However, for 80% and 90% identical point patterns the similarity values were close to the expected values and

the similarity test performs well. The standard deviations of these bandwidths are all quite similar except for bandwidth selection on a geometry window.

The similarity test yields larger than expected similarity values for the second simulation method. Abramson's adaptive bandwidth performs best in terms of mean (0.8455, 0.8616, 0.8800) and median (0.8640, 0.9062, 0.9151). It is still higher than expected when looking at 70% and 80% identical point patterns, but lower than the rest of the bandwidths. However, it has the largest standard deviation (0.1350, 0.1316, 0.1188) and coefficient of variation (0.1597, 0.1527, 0.1350). The reason for this might be that this bandwidth determines a bandwidth for each point in the spatial data set and a pixel image representation is obtained for each point. The rest of the bandwidths yields large similarity values (mean and median) with small standard deviations and coefficients of variation. Stoyan's rule of thumb yields the second closest mean (0.8790, 0.9172, 0.9606) and median (0.8733, 0.9160, 0.9611) to what is expected and small standard deviation (0.0699, 0.0517, 0.0794) and coefficient of variation (0.0794, 0.0564, 0.0267).

The third method yields higher similarity values than expected, particularly when considering the mean and median value. The standard deviation and coefficient of variation is small. Stoyan's rule of thumb yields the closest to expected similarity values where the means are 0.9092, 0.9333 and 0.9681, and the medians are 0.9229, 0.9464 and 0.9714. Note these values are still very high and might be because this case is highly theoretical.

Overall for the simulation study all bandwidths, except bandwidth selection based on a geometry window, perform quite well for the similarity test. Diggle's bandwidth performed best for the noisy patterns. Abramson's adaptive bandwidth performed best for point patterns that have uneven sample sizes. It can still be investigated how a change in constants for Abramson's adaptive bandwidth and Stoyan's rule of thumb influences the result of the similarity test as well as whether a change in the probability value f for bandwidth selection based on a geometry window will yield better results.

5. Conclusion

The robustness of the proposed spatial similarity test (Kirsten and Fabris-Rotelli, 2021) to different bandwidths was tested. Diggle's bandwidth (Diggle, 1985), likelihood cross-validation (Loader, 2006), Cronie & van Lieshout (Cronie and Van Lieshout, 2018), Scott's rule of thumb (Odell-Scott, 1992), Abramson's adaptive bandwidths (Abramson, 1982), bandwidth selection based on a geometry window (Baddeley et al., 2015) and Stoyan's rule of thumb (Stoyan and Stoyan, 1994) were the different bandwidths used to compute the pixel image representation in Step 1 of the spatial similarity test in order to test the robustness of the test. A suggestion for future work is to investigate how a change in constants for Abramson's adaptive bandwidth and Stoyan's rule of thumb influences the result of the similarity test as well as whether a change in the probability value f for bandwidth selection based on a geometry window will yield better results. Another suggestion for future work is a further investigation on the negative similarity values obtained.

The applications in Section 5 also provided a real data case for testing similarity across different windows. It was observed that different bandwidths perform differently for point patterns of different sizes and point patterns with different windows.

Table 3. Summary statistics of the results from the proposed spatial similarity test.

	Diggle	Likelihood cross	Cronie van Lieshout	Scott	Abramson	Geometry window	Stoyan
Method one							
Mean							
70%	0.7316	0.7782	0.7793	0.7941	0.7746	0.8973	0.6652
80%	0.8047	0.8365	0.8458	0.8525	0.8379	0.9283	0.7548
90%	0.8899	0.9062	0.9179	0.9244	0.9071	0.9686	0.8539
Median							
70%	0.7387	0.7819	0.7919	0.8119	0.7824	0.9330	0.6880
80%	0.8271	0.8670	0.8793	0.8888	0.8642	0.9572	0.7889
90%	0.9221	0.9453	0.9497	0.9582	0.9492	0.9872	0.8905
Standard deviation							
70%	0.1401	0.1585	0.1346	0.1373	0.1645	0.0980	0.1447
80%	0.1486	0.1519	0.1537	0.1513	0.1547	0.1083	0.1621
90%	0.1478	0.1246	0.1378	0.1307	0.1351	0.0690	0.1725
Coefficient of variation							
70%	0.1914	0.2037	0.1727	0.1729	0.2124	0.1092	0.2175
80%	0.1847	0.1815	0.1817	0.1775	0.1846	0.1167	0.2148
90%	0.1661	0.1375	0.1501	0.1414	0.1489	0.0712	0.2020
Method two							
Mean							
70%	0.9186	0.9421	0.9362	0.9424	0.8455	0.9749	0.8790
80%	0.9443	0.9616	0.9561	0.9598	0.8616	0.9839	0.9172
90%	0.9704	0.9799	0.9780	0.9806	0.8800	0.9930	0.9606
Median							
70%	0.9377	0.9616	0.9499	0.9516	0.8640	0.9856	0.8733
80%	0.9542	0.9752	0.9693	0.9691	0.9062	0.9925	0.9160
90%	0.9818	0.9877	0.9855	0.9858	0.9151	0.9972	0.9611
Standard deviation							
70%	0.0680	0.0601	0.0567	0.0465	0.1350	0.0320	0.0699
80%	0.0459	0.0405	0.0416	0.0383	0.1316	0.0228	0.0517
90%	0.0339	0.0230	0.0235	0.0193	0.1188	0.0089	0.0794
Coefficient of variation							
70%	0.0740	0.0638	0.0606	0.0493	0.1597	0.0328	0.0794
80%	0.0486	0.0421	0.0435	0.0399	0.1527	0.0232	0.0564
90%	0.0350	0.0235	0.0240	0.0197	0.1350	0.0089	0.0267
Method three							
Mean							
70%	0.9627	0.9567	0.9672	0.9682	0.9326	0.9842	0.9092
80%	0.9667	0.9640	0.9685	0.9734	0.9465	0.9874	0.9333
90%	0.9814	0.9827	0.9842	0.9867	0.9750	0.9926	0.9681
Median							
70%	0.9824	0.9697	0.9797	0.9800	0.9557	0.9913	0.9229
80%	0.9836	0.9752	0.9810	0.9856	0.9636	0.9927	0.9464
90%	0.9836	0.9892	0.9933	0.9938	0.9841	0.9968	0.9714
Standard deviation							
70%	0.0360	0.0360	0.0312	0.0313	0.0589	0.0174	0.0659
80%	0.0394	0.0307	0.0291	0.0258	0.0464	0.0127	0.0489
90%	0.0163	0.0155	0.0200	0.0150	0.0249	0.0106	0.0229
Coefficient of variation							
70%	0.0374	0.0377	0.0323	0.0324	0.0632	0.0177	0.0725
80%	0.0408	0.0319	0.0300	0.0265	0.0491	0.0129	0.0524
90%	0.0166	0.0158	0.0203	0.0152	0.0256	0.0107	0.0237

References

- ABRAMSON, I. S. (1982). On bandwidth variation in kernel estimates: A square root law. *The Annals of Statistics*, **10**, 1217–1223.
- ALBA-FERNÁNDEZ, M., ARIZA-LÓPEZ, F., JIMÉNEZ-GAMERO, M. D., AND RODRÍGUEZ-AVI, J. (2016). On the similarity analysis of spatial patterns. *Spatial Statistics*, **18**, 352–362.
- ANDRESEN, M. A. (2009). Testing for similarity in area-based spatial patterns: A nonparametric Monte Carlo approach. *Applied Geography*, **29**, 333–345.
- ANDRESEN, M. A. AND LINNING, S. J. (2012). The (in)appropriateness of aggregating across crime types. *Applied Geography*, **35**, 275–282.
- ANDRESEN, M. A. AND MALLESON, N. (2013a). Crime seasonality and its variations across space. *Applied Geography*, **43**, 25–35.
- ANDRESEN, M. A. AND MALLESON, N. (2013b). Spatial heterogeneity in crime analysis. *In Crime Modeling and Mapping Using Geospatial Technologies*. Springer, 3–23.
- ANDRESEN, M. A. AND MALLESON, N. (2014). Police foot patrol and crime displacement: A local analysis. *Journal of Contemporary Criminal Justice*, **30**, 186–199.
- BADDELEY, A., RUBAK, E., AND TURNER, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. CRC press.
- BERMAN, M. AND DIGGLE, P. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society: Series B (Methodological)*, **51**, 81–92.
- BORRAJO, M., GONZÁLEZ-MANTEIGA, W., AND MARTÍNEZ-MIRANDA, M. (2020). Testing for significant differences between two spatial patterns using covariates. *Spatial Statistics*, **40**, 100379.
- COX, D. R. AND ISHAM, V. (1980). *Point Processes*, volume 12. CRC Press.
- CRESSIE, N. (2015). *Statistics for Spatial Data*. John Wiley & Sons.
- CRONIE, O. AND VAN LIESHOUT, M. N. M. (2018). A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika*, **105**, 455–462.
- DIGGLE, P. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **34**, 138–147.
- GUAN, Y. (2007). A least-squares cross-validation bandwidth selection approach in pair correlation function estimations. *Statistics & Probability Letters*, **77**, 1722–1729.
- HALL, P. AND MARRON, J. (1988). Variable window width kernel estimates of probability densities. *Probability Theory and Related Fields*, **80**, 37–49.
- HEIDENREICH, N.-B., SCHINDLER, A., AND SPERLICH, S. (2013). Bandwidth selection for kernel density estimation: A review of fully automatic selectors. *ASIA Advances in Statistical Analysis*, **97**, 403–433.
- ILLIAN, J., PENTTINEN, A., STOYAN, H., AND STOYAN, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons.
- KIRSTEN, R. AND FABRIS-ROTELLI, I. N. (2021). A generic test for the similarity of spatial data. *South African Statistical Journal*, **55**, 55–71.
- KUTER, S., USUL, N., AND KUTER, N. (2011). Bandwidth determination for kernel density analysis

- of wildfire events at forest sub-district scale. *Ecological Modelling*, **222**, 3033–3040.
- LINNING, S. J. (2015). Crime seasonality and the micro-spatial patterns of property crime in Vancouver, BC and Ottawa, ON. *Journal of Criminal Justice*, **43**, 544–555.
- LOADER, C. (2006). *Local Regression and Likelihood*. Springer Science & Business Media.
- ODELL-SCOTT, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons Incorporated.
- STOYAN, D. AND STOYAN, H. (1994). *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics*, volume 302. John Wiley & Sons Incorporated.
- WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, **13**, 600–612.
- WEGLARCZYK, S. (2018). Kernel density estimation and its application. *In ITM Web of Conferences*, volume 23. EDP Sciences.