A Most Powerful Test for the Adequateness of an Asymptotic Spatial Regression Model when the Observation is Disturbed by the Set-Indexed Brownian Sheet

WAYAN SOMAYASA
Halu Oleo University
Department of Mathematics
Kendari, 93232, INDONESIA
wayan.somayasa@uho.ac.id

ASRUL SANI
Halu Oleo University
Department of Mathematics
Kendari, 93232, INDONESIA
saniasrul2001@yahoo.com

YULIUS BARA PASOLON
Halu Oleo University
Department of Agricultural Technology
Kendari, 93232, INDONESIA
yulpas_59@yahoo.com

Abstract: In this work we establish an optimal test for checking the appropriateness of a spatial regression model. In the study of model check for regression, the correctness of an assumed model is investigated by the partial sums of the residuals. In this work an inverted procedure is proposed in that we firstly embed the observation into a partial sums process to get the corresponding asymptotic regression model. Instead of considering the residuals of the model, we derive the Cameron-Martin density of the observation. For simple hypotheses under H_0 as well as under H_1 we derive the Neyman-Person test based on the ratio of the densities under H_0 and H_1. Interestingly, the rejection region can be exactly computed as an integral with respect to the partial sums process of the observation. An application of the procedure to a real data is also discussed.

Key–Words: Most powerful test, set-indexed Brownian sheet, Neyman-Pearson test, Cameron-Martin density, reproducing kernel Hilbert space

1 Introduction

The application of partial sums method in spatial regression has been studied in many literatures. MacNeill and Jandhyalla [16] and Xie and MacNeill [23] utilized respectively the ordinary and set-indexed partial sums of the spatial least squares residuals of polynomial regressions in detecting the existence of a boundary in the experimental region. They derived the limit process by adopting the technique proposed in MacNeill [14, 15]. Recently Somayasa and at al. [20, 21] developed asymptotic method in model check for spatial regression based on the set-indexed partial sums of the residuals. By considering an equally spaced experimental design (regular lattice) as the experimental design they obtained the limit process which is a functional of the set-indexed Brownian sheet by extending the geometric approach of Bischoff [7], Bischoff and Somayasa [9] and Somayasa [18] and by applying the existing uniform central limit theorem investigated in Alexander and Pyke [1] and Pyke [17].

In this paper we establish asymptotic model-check for spatial regression by proposing a different approach in that instead of considering the partial sums of the residuals we firstly attach the observations into the set-indexed stochastic process (random function). Next we derive Neyman-Pearson test procedure which is most powerful test based on the ratio of the density functions of the processes under the hypotheses. In contrast to the methods studied in the literatures mentioned above in this paper we use unequally spaced experimental design obtained by sampling the observations according to a probability measure, see Bischoff [6] and Somayasa [19]. So that from the practical view our proposed method seems to be more flexible in the sense of economic, technical or ecological reasons.

Let us consider a spatial regression model

$$Y(\mathbf{x}) = g(\mathbf{x}) + \varepsilon(\mathbf{x}),\ \mathbf{x} \in D \subset \mathcal{R}^d,$$

where g is an unknown function of bounded variation on the experimental region D and ε is unobserved random error with $\mathbf{E}(\varepsilon(\mathbf{x})) = 0$ and $Var(\varepsilon(\mathbf{x})) = \sigma^2 > 0$ for every $\mathbf{x} \in D$. In this paper we restrict the consideration to the experimental region given by a two dimensional rectangle $D := [a_1, a_2] \times [b_1, b_2]$, for $a_1 < a_2$ and $b_1 < b_2$. Result for higher dimensional rectangle can be obtained immediately. Given a probability measure P_0 on the Borel σ-algebra $\mathcal{B}(D)$ we construct the experimental design

$$\Xi_n := \{(t_{n\ell}, s_{nk}) :\ 1 \leq \ell, k \leq n\}$$

on D by performing a sampling procedure according to the method proposed in [19]. By this sampling scheme Ξ_n is not necessarily a regular lattice, unless P_0 is a uniform probability measures on $\mathcal{B}(D)$. Let P_n be a discrete probability measure on $\mathcal{B}(D)$ associated to Ξ_n defined by

$$P_n(B) := \frac{1}{n^2}\sum_{\ell=1}^{n}\sum_{k=1}^{n}\delta_{(t_{n\ell},s_{nk})}(B),\ B\in\mathcal{B}(D),$$

where for a fixed $(t_{n\ell},s_{nk})\in\Xi_n$, $\delta_{(t_{n\ell},s_{nk})}$ is the Dirac measure in the point $(t_{n\ell},s_{nk})$, see Bauer [4]. We note that P_n can also be equivalently written as

$$P_n(B) := \frac{1}{n^2}\sum_{\ell=1}^{n}\sum_{k=1}^{n}\mathbf{1}_B(t_{n\ell},s_{nk}), B\in\mathcal{B}(D),$$

where $\mathbf{1}_B$ is the indicator function of B. By this sampling scheme we get the property that $P_n \Rightarrow P_0$ as $n\to\infty$. Throughout the paper $\Rightarrow$ stands for the convergence in distribution in the sense of Billingsley [5].

Let $\mathbf{W}$ and $\mathbf{V}$ be finite dimensional spaces defined by $\mathbf{W} := [f_1,\ldots,f_p]$ and $\mathbf{V} := [f_1,\ldots,f_p,f_{p+1},\ldots,f_m]$, $p\le m$, where $f_1,\ldots,f_p,f_{p+1},\ldots,f_m$ are known regression functions which are assumed to be orthogonal as functions in $L_2(D,P_0)$, where $L_2(D,P_0)$ is the space of squared integrable functions on D with respect to P_o. Suppose g is decomposable as $g=g_1\oplus g_2$, where $g_1\in\mathbf{W}$ and $g_2\in\mathbf{V}\cap\mathbf{W}^C$, such that $\langle g_1,g_2\rangle_{L_2}=0$, where $\mathbf{W}^C$ is the complement of $\mathbf{W}$. It is the purpose of the present paper to develop an optimal test procedure for the hypotheses

$$H_0: g\in\mathbf{W}\ against\ H_1: g\in\mathbf{V} \tag{1}$$

so that upon testing the hypotheses it can be concluded whether $f_1,\ldots,f_p$ is adequate or we additionally need $f_{p+1},\ldots,f_m$ for representing g. By the assumption, testing (1) is equivalent to the problem of testing the following

$$H_0: g_2\equiv 0\ against\ H_1: g_2\equiv f, \tag{2}$$

for some function $f\in\mathbf{V}\cap\mathbf{W}^C$. Thus we consider under H_0 as well as H_1 simple hypotheses.

Let $Y(\Xi_n) := \{Y(t_{n\ell},s_{nk}):\ 1\le\ell,k\le n\}$ be the array of independent observations of Y over Ξ_n, such that

$$Y(\Xi_n) = g(\Xi_n)+\varepsilon(\Xi_n), \tag{3}$$

where $g(\Xi_n) := (g(t_{n\ell},s_{nk}))_{\ell=1,k=1}^{n,n}\in\mathcal{R}^{n\times n}$, and $\varepsilon(\Xi_n) := (\varepsilon(t_{n\ell},s_{nk})_{\ell=1,k=1}^{n,n}$ is an $n\times n$ matrix of random errors having independent and identically distributed components with $\mathbf{E}(\varepsilon(t_{n\ell},s_{nk}))=0$ and $Var(\varepsilon(t_{n\ell},s_{nk}))=\sigma^2>0$, for $1\le\ell,k\le n$. In the classical study an approach for testing (1) and (2) has been proposed by Arnold [2] and Arnold [3] by investigating the ratio between the length of the residual of the observation of Model 3 under H_0 and under H_1. In [19] the limit process of the ordinary partial sums process of the residual was investigated.

In this paper we embed the observations into a stochastic process using the set-indexed partial sums operator defined below. Beforehand we give a formal definition of set-indexed Brownian sheet.

Definition 1 *(Gaenssler [10]) Let $\mathcal{A}\subset\mathcal{B}(D)$ be a Vapnik-Chervonenkis class (VCC) of subsets of D. A pseudo metric d_{P_0} on $\mathcal{A}\times\mathcal{A}$ is defined by $d_{P_0}(A_1,A_2) := P_0(A_1\triangle A_2)$. Let $\ell^\infty(\mathcal{A})$ be a subset of $\mathcal{A}$ defined by*

$$\ell^\infty(\mathcal{A}) = \left\{w:\mathcal{A}\to\mathcal{R}|\ \|w\|_{\mathcal{A}} := \sup_{A\in\mathcal{A}}|w(A)|<\infty\right\}$$

Furthermore let $U^b(\mathcal{A},d_{P_0})$ be the space of functions in $\ell^\infty(\mathcal{A})$ that is d_{P_0}-uniformly continuous. A centered Gaussian process $\mathcal{W}_{P_0} := \{\mathcal{W}_{P_0}(A),\ A\in\mathcal{A}\}$ is called $\mathcal{A}$-indexed Brownian sheet (Gaussian white noise) with the control measure P_0 if and only if for every $A,B\in\mathcal{A}$, $\mathbf{E}(\mathcal{W}_{P_0}(A)\mathcal{W}_{P_0}(B)) = P_0(A\cap B)$. The sample paths of $\mathcal{W}_{P_0}$ are concentrated in $U^b(\mathcal{A},d_{P_0})$. The properties of $\mathcal{W}_{P_0}$ is summarized below:

1. *$Var(\mathcal{W}_{P_0}(A)) = P_0(A),\ \forall A\in\mathcal{A}$.*
2. *If $A_1,\ldots,A_n$ are disjoint, then the random variables $\mathcal{W}_{P_0}(A_1),\ldots,\mathcal{W}_{P_0}(A_n)$ are mutually independent.*
3. *If $A_1,\ldots,A_n$ are disjoint, then*

$$\sum_{j=1}^{n}\mathcal{W}_{P_0}(A_j) = \mathcal{W}_{P_0}\left(\bigcup_{j=1}^{n}A_j\right),\ a.s.$$

Definition 2 *An operator $\mathbf{S}_n:\mathcal{R}^{n\times n}\to\ell^\infty(\mathcal{A})$, defined by*

$$\mathbf{S}_n(\mathbf{A}_{n\times n})(B) := \frac{1}{n}\sum_{\ell=1}^{n}\sum_{k=1}^{n}\mathbf{1}_B(t_{n\ell},s_{nk})a_{\ell k},$$

for every $\mathbf{A}_{n\times n}=(a_{\ell k})_{k=1,\ell=1}^{n,n}\in\mathcal{R}^{n\times n}$ and $B\in\mathcal{A}$ is called $\mathcal{A}$-indexed partial sums operator. See also Gaenssler [10].

Theorem 3 *Let $\varepsilon(\Xi_n) := (\varepsilon(t_{n\ell},s_{nk}))_{\ell=1,k=1}^{n,n}$ in Model 3 consist of independent and identically distributed random variables with $\mathbf{E}(\varepsilon(t_{n\ell},s_{nk}))=0$ and $Var(\varepsilon(t_{n\ell},s_{nk}))=\sigma^2$, $n\ge1$. Then we have*

$$\frac{1}{\sigma}\mathbf{S}_n(\varepsilon(\Xi_n))(\cdot)\Rightarrow\mathcal{W}_{P_0}(\cdot),\ as\ n\to\infty.$$

Proof: We refer the reader to Gaenssler [10].

Based on the reason that $g \in \mathbf{W}$ is equivalent to $\frac{1}{n}g \in \mathbf{W}$ for all $n \geq 1$, we test the hypotheses by observing the localized version of (3) defined by

$$Y^{loc}(\Xi_n) = \frac{1}{n}g(\Xi_n) + \varepsilon(\Xi_n). \quad (4)$$

Then by the linearity of $\mathbf{S}_n$, we have

$$\mathbf{S}_n(Y^{loc}(\Xi_n)) = \mathbf{S}_n(\frac{1}{n}g(\Xi_n)) + \mathbf{S}_n(\varepsilon(\Xi_n)).$$

Since g has bounded variation on D it can be easily shown that $\frac{1}{\sigma}\mathbf{S}_n(\frac{1}{n}g(\Xi_n))$ converges uniformly to $\frac{1}{\sigma}\varphi_g \in \ell^\infty(\mathcal{A})$, where $\varphi_g(A) := \int_A g(t,s)P_0(dt,ds)$, for any $A \in \mathcal{A}$. Thus combining this result with Theorem 3 we get

$$\frac{1}{\sigma}\mathbf{S}_n(Y^{loc}(\Xi_n))(\cdot) \Rightarrow \mathcal{Y} := \frac{1}{\sigma}\varphi_g(\cdot) + \mathcal{W}_{P_0}(\cdot).$$

The last result shows that the partial sums of the localized model can be approximated by $\mathcal{Y}$ which is a signal plus noise model with the signal $\varphi_g(\cdot)$ and the noise $\mathcal{W}_{P_0}$. Hence the problem of testing the hypothesis $H_0 : g_2 \equiv 0$ can be handled by observing the asymptotic model $\mathcal{Y}$ and considering the hypothesis $H_0 : \varphi_{g_2} \equiv 0$, since $= \varphi_g(\cdot) = \varphi_{g_1}(\cdot) + \varphi_{g_2}(\cdot)$. This test problem will be the main objective of the current work. Without loss of generality we assume $\sigma^2 = 1$.

The rest of the present paper is organized as follows. In Section 2 we investigate the Cameron-Martin formula for the density of $\mathcal{Y}$. For that we adopt the technique proposed in Bischoff and Gegg [8] which is inspired by the work of Lifshit [13]. In Section 3 we establish the uniformly most powerful test for testing the hypotheses formulated above. Application of the test procedure will be presented in Section 4. We also present conclusion and remark for future work at the end of the paper.

2 Cameron-Martin Formula for the $\mathcal{A}$-Indexed Gasussian Sheet

First we derive the reproducing kernel Hilbert space (RKHS) of $\mathcal{W}_{P_0}$ which plays important role throughout the work. Let $\mathbf{J} : L_2(D,P_0) \to U^b(\mathcal{A}, d_{P_0})$ be a linear operator defined by $(\mathbf{J}\ell)(A) := \int_A \ell \; dP_0$. Then $\mathbf{J}$ is injective. The dual space of $U^b(\mathcal{A}, d_{P_0})$ is given by the space of signed measure on $\mathcal{A}$, denoted by $U^{b*}(\mathcal{A}, d_{P_0})$. The duality is defined by $\langle \mu, f\rangle := \int_D f \; d\mu$, for $(\mu, f) \in U^{b*}(\mathcal{A}, d_{P_0}) \times U^b(\mathcal{A}, d_{P_0})$. Next we define the operator $\mathbf{J}^* : U^{b*}(\mathcal{A}, d_{P_0}) \to L_2(D,P_0)$, defined by $(\mathbf{J}^*\mu)(x,y) = \mu([x,a_2] \times [y,b_2])$, for $(x,y) \in D$. Then for every $\mu \in U^{b*}(\mathcal{A}, d_{P_0})$ and $A \in \mathcal{A}$ we get

$$\begin{aligned}
(\mathbf{J}\mathbf{J}^*\mu)(A) &= \int_A (\mathbf{J}^*\mu)(x,y)P_0(dx,dy) \\
&= \int_A \mu([x,a_2] \times [y,b_2])P_0(dx,dy) \\
&= \int_D \int_D \mathbf{1}_{\{(x,y)\in A\}}\mathbf{1}_{\{x\leq s, y\leq t\}}P_0(dx,dy)\mu(ds,dt) \\
&= \int_D P_0(A \cap [a_1,s] \times [b_1,t])\mu(ds,dt) \\
&= (K\mu)(A),
\end{aligned}$$

where K is the covariance operator of the process $\mathcal{W}_{P_0}$. Thus the sufficient condition of Theorem 4.1 in Lifshits [13] is fulfilled. Therefore it can be concluded that the RKHS of $\mathcal{W}_{P_0}$ is given by

$$\mathcal{H}_{\mathcal{W}_{P_0}} = \left\{h : \; h(A) = \int_A \ell \, dP_0, \ell \in L_2(D,P_0)\right\}.$$

The space $\mathcal{H}_{\mathcal{W}_{P_0}}$ is furnished with the inner product and norm defined by

$$\langle h_1, h_2\rangle_{\mathcal{H}_{\mathcal{W}_{P_0}}} := \langle \ell_1, \ell_2\rangle_{L_2}, \; \|h\|_{\mathcal{H}_{\mathcal{W}_{P_0}}} = \|\ell\|_{L_2}.$$

Let P be the distribution of $\mathcal{W}_{P_0}$ on $U^b(\mathcal{A}, d_{P_0})$. For any function $h \in U^b(\mathcal{A}, d_{P_0})$, let P_h be the distribution of $h + \mathcal{W}_{P_0}$ on $U^b(\mathcal{A}, d_{P_0})$, where for every Borel set $B \subset U^b(\mathcal{A}, d_{P_0})$, $P_h(B) := P(B - h)$. The function h is called a shift. In the case $h \in \mathcal{H}_{\mathcal{W}_{P_0}}$ then h is called admissible shift.

Now we are ready to state the Cameron-Martin theorem for the shifted Gaussian process $h + \mathcal{W}_{P_0}$.

Theorem 4 *P_h is absolutely continuous with respect to P on $U^b(\mathcal{A}, d_{P_0})$ if and only if $h \in \mathcal{H}_{\mathcal{W}_{P_0}}$. If $h \in \mathcal{H}_{\mathcal{W}_{P_0}}$ with $h(A) = \int_A \ell \, dP_0$ for $A \in \mathcal{A}$, then*

$$\frac{dP_h}{dP}(w) = \exp\left\{\int_D \ell \; dw - \frac{1}{2}\|h\|^2_{\mathcal{H}_{\mathcal{W}_{P_0}}}\right\}, \quad (5)$$

for P-almost all $w \in U^b(\mathcal{A}, d_{P_0})$, where the integral $\int_D \ell \; dw$ is in the sense of Wiener integral defined in appendix.

Proof. By recalling Theorem 5.1 of [13] we get the general formula for the density of P_h with respect to P, that is

$$\frac{dP_h}{dP}(w) = \exp\left\{z(w) - \frac{1}{2}\|h\|^2_{\mathcal{H}_{\mathcal{W}_{P_0}}}\right\},$$

for P-almost all $w \in U^b(\mathcal{A}, d_{P_0})$, where z is a linear measurable functional on $U^b(\mathcal{A}, d_{P_0})$ such that $\mathbf{I}z = h$. Comparing between this general formula and Equation 5 we only need to show that $(\mathbf{I}z)(A) = h(A) =$

$\int_A \ell\, dP_0$ for every $A \in \mathcal{A}$. Interested reader is referred to [13], pp. 22–37 for a concise discussion regarding the operator $\mathbf{I}$. We claim that $z(w) = \int_D \ell\, dw$. Then for every $A \in \mathcal{A}$ we have

$$\begin{aligned}(\mathbf{I}z)(A) &= \delta_A(\mathbf{I}z) = \langle \delta_A, \mathbf{I}z\rangle = \langle \mathbf{I}^*\delta_A, z\rangle \\ &= \mathbf{E}(W_{P_0}(A)z(W_{P_0})) \\ &= \mathbf{E}\left(\int_D \mathbf{1}_A\, dW_{P_0} \int_D \ell\, dW_{P_0}\right) \\ &= \int_A \ell\, dP_0 = h(A).\end{aligned}$$

Thus we show that our claim is correct and therefore the proof in complete. Clearly φ_{g_1} and φ_f are in $\mathcal{H}_{\mathcal{W}_{P_0}}$ having the $L_2(D, P_0)$-density g_1 and f, respectively.

3 Uniformly Most Powerful Test

We consider for the moment the simple hypotheses $H_0 : \varphi_{g_2} \equiv 0$ against $H_1 : \varphi_{g_2} \equiv \varphi_{f_0}$, for some $f_0 \in \mathbf{V} \cap \mathbf{W}^C$. Under the situation of H_0 the sample $\mathcal{Y}(\cdot)$ is generated from the model $\mathcal{Y}(\cdot) = \varphi_{g_1}(\cdot) + \mathcal{W}_{P_0}(\cdot)$, whereas under H_1, $\mathcal{Y}(\cdot) = \varphi_{g_1+f_0}(\cdot) + \mathcal{W}_{P_0}(\cdot)$. The following theorem presents a most powerful test of size α for testing the simple hypotheses. A similar result was obtained in the work of Gegg [11] for Brownian motion on $[0, 1]$.

Theorem 5 *Observing the model $\mathcal{Y} = \varphi_g(\cdot)+\mathcal{W}(\cdot)$, a uniformly most powerful test of size α for the hypotheses $H_0 : \varphi_{g_2} \equiv 0$ against $H_1 : \varphi_{g_2} \equiv f_0$ for some $f_0 \in \mathbf{V} \cap \mathbf{W}^C$ will reject H_0 if and only if*

$$\int_D f_0 d\mathcal{Y} \geq z_{1-\alpha}\|f_0\|_{L_2}.$$

where $z_{1-\alpha}$ is the $(1-\alpha)$ quantile of the standard normal distribution.

Proof. Let ψ_0 and ψ_1 be the density of $\mathcal{Y}$ under H_0 and H_1, respectively. Then by formula (5) we have

$$\begin{aligned}\frac{\psi_0(\mathcal{Y})}{\psi_1(\mathcal{Y})} &= \frac{\exp\left\{-\frac{1}{2}\|\varphi_{g_1}\|^2_{\mathcal{H}_{\mathcal{W}_{P_0}}} + \int_D g_1 d\mathcal{Y}\right\}}{\exp\left\{-\frac{1}{2}\|\varphi_{g_1+f_0}\|^2_{\mathcal{H}_{\mathcal{W}_{P_0}}} + \int_D (g_1+f_0) d\mathcal{Y}\right\}} \\ &= \exp\left\{\frac{1}{2}\|\varphi_{f_0}\|^2_{\mathcal{H}_{\mathcal{W}_{P_0}}} - \int_D f_0 d\mathcal{Y}\right\} \\ &= \exp\left\{\frac{1}{2}\|f_0\|^2_{L_2} - \int_D f_0\, d\mathcal{Y}\right\}.\end{aligned}$$

Here $\|\varphi_{g_1+f_0}\|^2_{\mathcal{H}_{\mathcal{W}_{P_0}}} = \|\varphi_{g_1}\|^2_{\mathcal{H}_{\mathcal{W}_{P_0}}} + \|f_0\|^2_{\mathcal{H}_{\mathcal{W}_{P_0}}}$ and $\langle g_1, f_0\rangle_{L_2} = 0$ by the orthogonality between g_1 and f_0. Furthermore, for $\mathcal{Y}(\cdot) = \varphi_{g_1} + \mathcal{W}_{P_0}$ it holds

$$\int_D f_0 d\mathcal{Y} = \langle f_0, g_1\rangle_{L_2} + \int_D f_0 d\mathcal{W}_{P_0}.$$

Hence for a predetermined α and a constant k we get

$$\begin{aligned}&P\left\{\frac{\psi_0(\mathcal{Y})}{\psi_1(\mathcal{Y})} \leq k|H_0\right\} = \alpha \\ \Leftrightarrow\ & P\left(\exp\left\{\frac{1}{2}\|\varphi_{f_0}\|^2_{\mathcal{H}_{\mathcal{W}_{P_0}}} - \int_D f_0 d\mathcal{Y}\right\} \leq k\right) = \alpha \quad \Leftrightarrow \\ & P\left\{\frac{\|f_0\|^2_{L_2}}{2} - \langle f_0, g_1\rangle_{L_2} - \int_D f_0\, d\mathcal{W}_{P_0} \leq lnk\right\} = \alpha \\ \Leftrightarrow\ & P\left\{\int_D f_0\, d\mathcal{W}_{P_0} \geq \frac{1}{2}\|f_0\|^2_{L_2} - lnk\right\} = \alpha \\ \Leftrightarrow\ & P\left\{\frac{1}{\|f_0\|_{L_2}}\int_D f_0\, d\mathcal{W}_{P_0} \geq k^*\right\} = \alpha,\end{aligned}$$

where $k^* := \frac{1}{2}\|f_0\|_{L_2} - \frac{lnk}{\|f_0\|_{L_2}}$. Moreover, since $\frac{1}{\|f_0\|_{L_2}}\int_D f_0\ d\mathcal{W}_{P_0} \sim N(0, 1)$, then we chose for $k^* = z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$-th quantile of the standard normal distribution. Hence we get

$$k = \exp\left\{\frac{1}{2}\|f_0\|^2_{L_2} - z_{1-\alpha}\|f_0\|_{L_2}\right\}.$$

Thus by the Neyman-Pearson theorem (cf. Theorem 3.2.1 in Lehmann and Romano [12]) a uniformly most powerful test of size α for the hypotheses $H_0 : \varphi_{g_2} \equiv 0$ against $H_1 : \varphi_{g_2} \equiv f_0$ for some $f_0 \in \mathbf{V} \cap \mathbf{W}^C$ will reject H_0 if and only if

$$T := \exp\left\{\frac{1}{2}\|f_0\|^2_{L_2} - \int_D f_0 d\mathcal{Y}\right\} \leq k,$$

for the constant k defined above. Equivalently, H_0 is rejected at level α if and only if

$$\int_D f_0 d\mathcal{Y} \geq z_{1-\alpha}\|f_0\|_{L_2}.$$

The following is the algorithm for conducting a model check in asymptotic regression involving the $\mathcal{A}$-indexed Brownian sheet as the noise:

1. Transform the the observation into the partial sums process $\mathcal{Y}$ by using the operator $\mathbf{S}_n$.
2. Compute the constant k.
3. Compute the test statistic T.
4. Draw decision by comparing T and k. Reject H_0 if $T \leq k$.

As an example we consider a first order model $Y(t, s) = \beta_0 + \beta_1 t + \beta_2 s + \varepsilon(t, s)$, for $(t, s) \in \mathbf{I}^2$. For technical reason suppose P_0 is the uniform probability measure on $\mathcal{B}(\mathbf{I}^2)$. By assuming that a constant model is adequate we admit a decomposition

$g(t,s) = g_1(t,s) \oplus g_2(t,s)$, with $g_1(t,s) = \beta_0$ and $g_2(t,s) = \beta_1 t + \beta_2 s$, for $(t,s) \in \mathbf{I}^2$. Hence the hypotheses of interest are $H_0 : g_2 \equiv 0$ against $H_1 : g_2(t,s) = f_0(t,s) = \beta_1 t + \beta_2 s$. By a little computation we get

$$\|f_0\|_{L_2} = \sqrt{\frac{\beta_1^2}{6} + \frac{\beta_1\beta_2}{4} + \frac{\beta_2^2}{6}}$$

and

$$\int_{\mathbf{I}^2} f_0 \, d\mathcal{Y} = \int_{\mathbf{I}^2} (\beta_1 t + \beta_2 s) d\mathcal{Y}(t,s).$$

4 Evaluating the Power

The power function of a the test derived above is defined as the probability of rejection of H_0 evaluated at any $f \in \mathbf{V} \cap \mathbf{W}^C$. Let us denote this function by $\Psi : \mathbf{V} \cap \mathbf{W}^C \to [0,1]$ defined by

$$\Psi(f) := \mathbf{P}\left\{ \frac{1}{\|f_0\|_{L_2}} \int_D f_0 d\mathcal{Y} \geq z_{1-\alpha} | \, \varphi_{g_2} \equiv \varphi_f \right\}.$$

If $\varphi_{g_2} \equiv \varphi_f$, then the sample is generated from the model

$$\mathcal{Y}(\cdot) = \varphi_{g_1}(\cdot) + \varphi_f(\cdot) + \mathcal{W}_{P_0}(\cdot).$$

Hence, we further have

$$\begin{aligned}
\frac{1}{\|f_0\|_{L_2}} \int_D f_0 d\mathcal{Y} &= \frac{1}{\|f_0\|_{L_2}} \int_D f_0 d(\varphi_{g_1} + \varphi_f + \mathcal{W}_{P_0}) \\
&= \frac{1}{\|f_0\|_{L_2}} \left(\int_D f_0 d\varphi_{g_1} + \int_D f_0 d\varphi_f + \int_D f_0 d\mathcal{W}_{P_0} \right) \\
&= \frac{1}{\|f_0\|_{L_2}} \left(\langle f_0, g_1 \rangle_{L_2} + \langle f_0, f \rangle_{L_2} + \int_D f_0 d\mathcal{W}_{P_0} \right) \\
&= \frac{1}{\|f_0\|_{L_2}} \left(\langle f_0, f \rangle_{L_2} + \int_D f_0 d\mathcal{W}_{P_0} \right).
\end{aligned}$$

Then the value of Ψ evaluated at any f is given by

$$\begin{aligned}
\Psi(f) := \mathbf{P}\left\{ \frac{1}{\|f_0\|_{L_2}} \int_D f_0 d\mathcal{W}_{P_0} \geq z_{1-\alpha} - \frac{\langle f_0, f \rangle_{L_2}}{\|f_0\|_{L_2}} \right\} \\
= 1 - \Phi\left(z_{1-\alpha} - \frac{\langle f_0, f \rangle_{L_2}}{\|f_0\|_{L_2}} \right),
\end{aligned}$$

where Φ is the probability distribution function of the standard normal distribution.

We notice that under H_0, that is when $f \equiv 0$, we have $\Psi(0) = 1 - \Phi(z_{1-\alpha}) = \alpha$ which is the size of the test. Conversely, since $\frac{\langle f_0, f \rangle_{L_2}}{\|f_0\|_{L_2}} > 0$ for any $f \in \mathbf{V} \cap \mathbf{W}^C$ with $f \neq 0$, we get $\Psi(f) \geq \alpha$. this means that the test maximizes the power under alternatives.

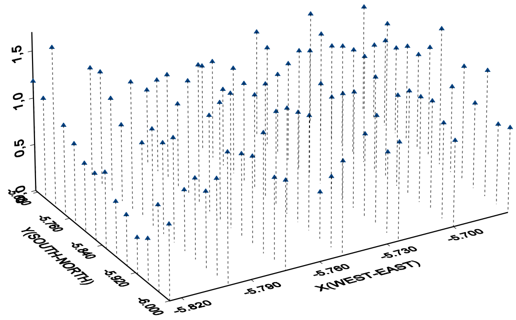

Figure 1: The scatter plot of the percentage of Ni with respect to the position of bores

.

5 Application

In this section we study the application of the proposed method to a mining data obtained from a mining company PT. ANEKA TAMBANG Tbk. in Pomalaa, Southeast Sulawesi, which is available also in Tahir [22]. The data consists of the percentage of nickel (Ni) observed over 14×7 lattice positions of bores on the exploration region with 14 equidistance rows running from south to north, and 7 equidistance column running from west to east. The measurements of the percentage of Ni are regarded as a realization of 14×7 dimensional matrix of independent response variables whose three dimensional scatter plot is presented in Figure 1 which shows more curvatures instead of planar. By this reason an adequate model for describing the relations between the position on the exploration region and the percentage of Ni is the second-order model.

Model check based on the partial sums of the residuals of this observation has been conducted in the work of [20, 21] which have been leading to the conclusion that a second-order model was significant for Ni data. Similarly we also consider the same hypotheses in the present paper. That is we test $H_0 : g_2 \equiv 0$ against $H_1 : g_2 \equiv f_0$, where $f_0(t,s) = t^2 s + t s^2$, for $(t,s) \in \mathbf{I}^2$. We get

$$\|f_0\|_{L_2} = \sqrt{\int_{\mathbf{I}^2} (t^2 s + t s^2)^2 dt ds} = \sqrt{31/120}$$

For computational reason we consider in this example the indexes $\{[0,t] \times [0,s] : \; 0 \leq t, s \leq 1\}$ instead of the family of all convex sets. Hence $\mathcal{Y}(t,s)$ is approximated by the partial sums defined in [9], [18], [19],

[20], [21] and [23]. That is

$$\begin{aligned}\mathcal{Y}(t,s) &\approx \mathbf{S}_n(Y(\Xi_{n_1\times n_2}))([n_1 t]/n_1, [n_2 s]/n_2)\\ &= \frac{1}{\sqrt{n_1 n_2}}\sum_{\ell=1}^{[n_1 t]}\sum_{k=1}^{[n_2 s]} Y_{\ell k},\end{aligned}$$

where $Y_{\ell k}$is the percentage of Ni on the point $(\ell/14, k/7)$, for $1 \le \ell \le 14$ and $1 \le k \le 7$. The integral $\int_{\mathbf{I}^2} f_0(t,s)\; d\mathcal{Y}(t,s)$ is approximated by the Riemann-Stieltjes sum, given by

$$\begin{aligned}&\int_{\mathbf{I}^2} f_0(t,s)\; d\mathcal{Y}(t,s)\\ &\approx \sum_{\ell=1}^{14}\sum_{k=1}^{7} f_0(\ell/14, k/7)\Delta_{\ell k}\mathbf{S}_n(Y(\Xi_{n_1\times n_2})),\end{aligned}$$

where

$$\begin{aligned}\Delta_{\ell k}\mathbf{S}_n(Y_{n_1\times n_2}) := &\mathbf{S}_n(Y_{n_1\times n_2})(\ell/n_1, k/n_2)\\ &-\mathbf{S}_n(Y_{n_1\times n_2})((\ell-1)/n_1, k/n_2)\\ &-\mathbf{S}_n(Y_{n_1\times n_2})(\ell/n_1, (k-1)/n_2)\\ &+\mathbf{S}_n(Y_{n_1\times n_2})((\ell-1)/n_1, (k-1)/n_2).\end{aligned}$$

After a set of computations using computer software R version 3.1.2, we get the approximated value of $\int_{\mathbf{I}^2} f_0(t,s) d\mathcal{Y}(t,s) = 0.80100$. Hence we obtain

$$\frac{1}{\|f_0\|_{L_2}}\int_{\mathbf{I}^2} f_0(t,s) d\mathcal{Y}(t,s) = \frac{0.80100}{\sqrt{31/120}} = 1.575949.$$

Since for $\alpha = 0.05$ the value of $z_{0.95} = 1.64485$, then we lead to the conclusion of the acceptance of H_0. Thus a similar result as in the computation conducted by [20] and [21] appears for the Nickel data.

6 Conclusion

The Cameron-Martin translation formula can be applied in obtaining the Neyman-Pearson test in model check for regression. Based on this approach an exact formula for constructing the size α rejection region is obtained.

Acknowledgements: The authors wish to thank Professor Wolgang Bischoff for the kind discussion during the preparation of the manuscript. We are grateful to RISTEK-DIKTI for the financial support.

Appendix A

A.1. Integral with Respect to $\mathcal{W}_{P_0}$

Definition A.1. (Lifshits [13]) Let $f := \sum_{j=1}^m a_j \mathbf{1}_{A_j}$ be any step function in $L_2(D, P_0)$. The Wiener integral of f with respect to $\mathcal{W}_{P_0}$ is defined by

$$\int_D f\; d\mathcal{W}_{P_0} = \sum_{j=1}^m c_j \mathcal{W}_{P_0}(A_j).$$

Proposition A.2. 1. For any $f_1 := \sum_{j=1}^m a_j \mathbf{1}_{A_j}$ and $f_2 := \sum_{j=1}^n c_j \mathbf{1}_{B_j}$, if $f_1 = f_2$, then

$$\sum_{j=1}^m a_j \mathcal{W}_{P_0}(A_j) = \sum_{j=1}^n c_j \mathcal{W}_{P_0}(B_j).$$

2. In the class of step functions it holds

$$\int_D (cf + bg)\; d\mathcal{W}_{P_0} = c\int_D f\; d\mathcal{W}_{P_0} + b\int_D g\; d\mathcal{W}_{P_0}.$$

3. For any step functions f and g it holds

$$\mathbf{E}\left(\int_D f\; d\mathcal{W}_{P_0}\int_D g\; d\mathcal{W}_{P_0}\right) = \int_D fg\; dP_0.$$

In particular $\int_D f\; d\mathcal{W}_{P_0} \sim N(0, \|f\|^2_{L_2(D,P_0)})$.

Definition A.3. (Lifshits [13]) For any $f \in L_2(D, P_0)$, $\int_D f\; d\mathcal{W}_{P_0} := \lim_{n\to\infty}\int_D f_n\; d\mathcal{W}_{P_0}$, for a sequence $(f_n)_{n\ge 1}$ of step functions that converges to f.

Remark A.4. Definition A.3 is well defined since the class of step functions is dense in the space $L_2(D, P_0)$. The limit exists and it does not depend on the choice of the sequence $(f_n)_{n\ge 1}$.

Proposition A.5. 1. For any $f_1, f_2 \in L_2(D, P_0)$, if $f_1 = f_2$, then

$$\int_E f_1\; d\mathcal{W}_{P_0} = \int_E f_2\; d\mathcal{W}_{P_0}.$$

2. For any $f, g \in L_2(D, P_0)$ and constants c, b it holds

$$\int_D (cf + bg)\; d\mathcal{W}_{P_0} = c\int_D f\; d\mathcal{W}_{P_0} + b\int_D g\; d\mathcal{W}_{P_0}.$$

3. For any $f, g \in L_2(D, P_0)$ we have

$$\mathbf{E}\left(\int_D f\; d\mathcal{W}_{P_0}\int_D g\; d\mathcal{W}_{P_0}\right) = \int_D fg\; dP_0.$$

References:

[1] K.S. Alexander and R. Pyke, A uniform central limit theorem for set-indexed partial-sum processes with finite variance, *The Annals of Probability*, 14, 1986, pp. 582–597.

[2] S.F. Arnold, Asymptotic validity of F tests for odinary linear model and the multiple correlation model, *Journal of the American Statistical Association*, 75 (372), 1980, pp. 890–894.

[3] S.F. Arnold, *The Theory of Linear Models and Multivarite Analysis*, John Wiley & Sons, Inc., New York, 1981.

[4] H. Bauer, *Measure and Integration Theory*, Walter de Gruyter, Berlin, 2001.

[5] P. Billingsley, *Convergence of Probability Measures (2nd Edition)*, John Wiley & Sons, Inc., New York, 1999.

[6] W. Bischoff, A functional central limit theorem for regression models, *Ann. Stat.* 6, 1998, pp. 1398–1410.

[7] W. Bischoff, The structure of residual partial sums limit processes of linear regression models, *Theory of Stochastic Processes*, 2, 2002, pp. 23–28.

[8] W. Bischoff and A. Gegg, The Cameron-Martin theorem for $(p-)$Slepian processes, Preprint, Catholic University Eichstaett-Ingolstadt,Germany, 2014.

[9] W. Bischoff and W. Somayasa, The limit of the partial sums process of spatial least squares residuals, *J. Multivariate Analysis*, 100, 2009, pp. 2167–2177.

[10] P.Gaenssler, On recent development in the theory of set-indexed processes (A unified approach to empirical and partial-sum processes) in *Asymptotic Statistics*, Springer, Berlin, 1993.

[11] A. Gegg, *Moving Windows zum Testen auf Change-Points (Sequentielle und beste Tests)*, Ph.D. Dissertation, Catholic University Eichstaett-Ingolstadt, Germany, 2013.

[12] E.L. Lehmann and J.P. Romano, *Testing Statitical Hypotheses*, 3rd edn., Springer, New York, 2005.

[13] M. Lifshits, *Lectures on Gaussian Processes*, Springer Briefs in Mathematics, Springer, Berlin, 2012.

[14] I.B. MacNeill, Properties of partial sums of polynomial regression residuals with applications to test for change of regression at unknown times, *Ann. Statist.* 6, 1978, pp. 422–433.

[15] I.B. MacNeill, Limit processes for sequences partial sums of regression residuals, *Ann. Probab.* 6, 1978, pp. 695–698.

[16] I.B. MacNeill and V.K. Jandhyala, *Change-point methods for spatial data, Multivariate Environmental Statistics eds. by G.P. Patil and C.R. Rao*, Elsevier Science Publishers B.V., 1993, pp. 298–306.

[17] W.J. Pyke, A uniform central limit theorem for partial sum processes indexed by sets, *Ann.Probab.* 79, 1983, pp. 219–240.

[18] W. Somayasa, On set-indexed residual partial sum limit process of spatial linear regression models, *J. Indones. Math. Soc.* **17(2)**, 2011, pp. 73–83.

[19] W. Somayasa, The partial sums of the least squares residuals of spatial observations sampled according to a probability measure", *J. Indones. Math. Soc.* **19(1)**, 2013, pp. 23–40.

[20] W. Somayasa, Asymptotic statistical model building based on the partial sums of the residuals of the observations with an application to mining industry, *Proceedings of the* 2^{nd} *International Conference on Mathematical, Computational and Statistical Sciences (MCSS)*, 2014, pp. 136–145.

[21] W. Somayasa, Ruslan, E. Cahyono, L.O. Engkoimani, Cheking adequateness of spatial regressions using set-indexed partial sums technique, *Fareast Journal of Mathematical Sciences*, 96(8), 2015, pp. 933–966.

[22] M. Tahir, Prediction of the amount of nickel deposit based on the results of drilling bores on several points (case study: south mining region of PT. Aneka Tambang Tbk., Pomalaa, Southeast Sulawesi), research report, Halu Oleo University, Kendari, 2010.

[23] L. Xie and I.B. MacNeill, Spatial residual processes and boundary detection, *South African Statist. J.* 4, 2006, pp. 33–53.