J. Indones. Math. Soc.
Vol. 17, No. 2 (2011), pp. 73–83.

# ON SET-INDEXED RESIDUAL PARTIAL SUM LIMIT PROCESS OF SPATIAL LINEAR REGRESSION MODELS

Wayan Somayasa

**Department of Mathematics, Universitas Haluoleo**
**Jl. H.E.A. Mokodompit, Kendari 93232, Indonesia**
**wayan.somayasa@yahoo.de**

**Abstract.** In this paper we derive the limit process of the sequence of set-indexed least-squares residual partial sum processes of observations obtained form a spatial linear regression model. For the proof of the result we apply the uniform central limit theorem of Alexander and Pyke [1] and generalize the geometrical approach of Bischoff [7] and Bischoff and Somayasa [8]. It is shown that the limit process is a projection of the set-indexed Brownian sheet onto the reproducing kernel Hilbert space of this process. For that we define the projection via Choquet integral [14, 15, 17] of the regression function with respect to the set-indexed Brownian sheet.

*Key words*: Set-indexed Brownian sheet, set-indexed partial sum process, spatial linear regression model, least-squares residual, Choquet integral.

**Abstrak.** Dalam makalah ini diturunkan proses limit dari barisan himpunan berindeks proses jumlah parsial residual least-square dari pengamatan yang diperoleh dari suatu model regresi linear spasial. Pembuktian hasil menggunakan teorema limit pusat dari Alexander dan Pyke [1] dan perumuman pendekatan geometris dari Bischoff [7] serta Bischoff dan Somayasa [8]. Hasil penelitian menunjukkan bahwa proses limit adalah sebuah proyeksi dari himpunan berindeks Brownian pada reproduksi kernel Ruang Hilbert dari proses ini. Untuk memperoleh hal tersebut didefinisikan proyeksi melalui integral Choquet [14, 15, 17] dari fungsi regresi terhadap himpunan berindeks Brownian.

*Kata kunci*: Himpunan berindeks Brownian, himpunan berindeks proses jumlah parsial, model regresi linear spasial, least-squares residual, integral Choquet.

---

*2000 Mathematics Subject Classification*: 62G, 62H, 62J.
Received: 07-04-2011, revised: 18-07-2011, accepted: 09-08-2011.

## 1. Introduction

Boundary detection problem (BDP) in spatial linear regression model (SLRM) is commonly handled by investigating the partial sum process of the least-squares residuals. The existence of the boundary is detected by conducting a nonparametric test based on the limit of the processes as for instance a test of Kolmogorov, Kolmogorov-Smirnov and Cramér-von Mises types. This technique can also be used to conduct model-check concerning the correctness of the assumed SLRM. MacNeill and Jandhyala [13] studied this residual partial sums process (RPSP) in order to detect the existence of a boundary in the experimental region, whereas in [8] it was studied from the perspective of model-check in SLRM.

The study has been extended to the set-indexed RPSP. To explain the problem in detail, suppose that a regression model

$$Y(\ell/n, k/n) = g(\ell/n, k/n) + \varepsilon(\ell/n, k/n)$$

is observed under an experimental design given by a regular lattice

$$\Xi_n := \{(\ell/n, k/n) : \ 1 \leq \ell, k \leq n\}, \ n \geq 1$$

on $I := [0,1] \times [0,1] \subset \mathbb{R}^2$, where $Y(\ell/n, k/n)$ and $\varepsilon(\ell/n, k/n)$ are the observation and the random error at the point $(\ell/n, k/n)$, respectively and $g$ is the unknown regression function defined on $I$. Throughout this paper for any real-valued function $f$ on $I$, the matrix $(f(\ell/n, k/n))_{k=1,\ell=1}^{n,n}$ will be denoted by $f(\Xi_n)$. So the whole observation can be presented as $Y_n = g(\Xi_n) + \mathbf{E}_n$, where $\mathbf{Y}_n := (Y_{\ell k})_{k=1,\ell=1}^{n}$ is the matrix of observations, and $\mathbf{E}_n := (\varepsilon_{\ell k})_{k=1,\ell=1}^{n,n}$ is the matrix of independent and identically distributed random errors with mean 0 and variance $\sigma^2$, $0 < \sigma^2 < \infty$. For a fixed $n \geq 1$, let $\mathbf{W}_n := [f_1(\Xi_n), \ldots, f_p(\Xi_n)]$ be a subspace of $\mathbb{R}^{n \times n}$ generated by $f_1(\Xi_n), \ldots, f_p(\Xi_n)$ , where $f_1, \ldots, f_p$ are known, linearly independent, real-valued regression functions defined on $I$. Model-check for this SLRM concerns with the question whether the hypothesis $H_0 : g(\Xi_n) \in \mathbf{W}_n$ is or is not supported by the sample. Let $\mathbf{P}_{\mathbf{W}_n}$ and $\mathbf{P}_{\mathbf{W}_n^\perp} = Id - \mathbf{P}_{\mathbf{W}_n}$ denote the orthogonal projector onto the subspace $\mathbf{W}_n$ and onto the orthogonal complement $\mathbf{W}_n^\perp$ of $\mathbf{W}_n$ with respect to the Euclidean inner product $\langle \cdot, \cdot \rangle_{\mathbb{R}^{n \times n}}$ on the Euclidean space of $n \times n$ matrices $\mathbb{R}^{n \times n}$, defined by $\langle \mathbf{A}, \mathbf{B} \rangle_{\mathbb{R}^{n \times n}} := trace(\mathbf{A}^\top \mathbf{B}), \ \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$. Analogous to the result in Arnold [2], the corresponding matrix of least squares residuals of the observations under $H_0$ is given by

$$\mathbf{R}_n := (r_{\ell k})_{k=1,\ell=1}^{n,n} = \mathbf{P}_{\mathbf{W}_n^\perp} \mathbf{Y}_n = \mathbf{P}_{\mathbf{W}_n^\perp} \mathbf{E}_n = \mathbf{E}_n - \mathbf{P}_{\mathbf{W}_n} \mathbf{E}_n. \tag{1}$$

The hypothesis mentioned above is commonly tested by using either the Cramér-von Mises or Kolmogorov-Smirnov functional of the set-indexed RPSP

$$Z_n := \left\{ \sum_{k=1}^{n} \sum_{\ell=1}^{n} n r_{\ell k} \lambda_I (B \cap C_{\ell k}) : \ B \in \mathcal{A} \right\},$$

indexed by the Borel $\sigma-$algebra $\mathcal{A}$ over $I$, where $\lambda_I$ is the corresponding Lebesgue measure on $\mathcal{A}$, and $C_{\ell k}$ is the half-open rectangle $((\ell-1)/n, \ell/n] \times ((k-1)/n, k/n]$.

Xie and MacNeill [18] established the limit process of the sequence $(Z_n)_{n\geq 1}$ by applying the method proposed by MacNeill [12]. Their result were presented as a complicated function of the set-indexed Brownian sheet. The aim of this paper is to derive the limit of the set-indexed RPSP by applying the geometrical approach of [7, 8] which is different form that of [13, 18], see Section 2. By our consideration we get the structure of the limit process as a projection of the set-indexed Brownian sheet onto a certain subspace of its reproducing kernel Hilbert space (RKHS). Under the alternative hypotheses we will observe localized nonparametric regression models $Y_n = g^{loc}(\Xi_n) + \mathbf{E}_n$, where $g^{loc}(\Xi_n) := \frac{1}{n}g(\Xi_n)$. Several examples are discussed in Section 3.

## 2. Main Results

Let $\mathcal{C}(A)$ be the space of bounded, real-valued set functions which are continuous on $\mathcal{A}$ in the symmetric difference pseudometric $d_{\lambda_I}$, see [1, 4, 10]. As usual $\mathcal{C}(\mathcal{A})$ is furnished with the norm $\|\cdot\|_{\mathcal{A}}$, defined by $\|u\|_{\mathcal{A}} := \sup_{A\in\mathcal{A}}|u(A)|$, $A \in \mathcal{C}(\mathcal{A})$, and every function $\nu \in \mathcal{C}(\mathcal{A})$ is further assumed to satisfies $\nu(\emptyset) = 0$. A set-indexed partial sum operator is a linear operator $\mathbf{T}_n : \mathbb{R}^{n\times n} \to \mathcal{C}(\mathcal{A})$, such that for every $\mathbf{A}_n := (a_{\ell k})_{k=1,\ell=1}^{n,n} \in \mathbb{R}^{n\times n}$, and $B \in \mathcal{A}$,

$$\mathbf{T}_n(\mathbf{A}_n)(B) := \sum_{k=1}^{n}\sum_{\ell=1}^{n} n a_{\ell k}\lambda_I(B \cap C_{\ell k}),\ \mathbf{T}_n(\mathbf{A}_n)(\emptyset) := 0.$$

Thus the set-indexed RPSP $Z_n$ is a stochastic process with sample path in $\mathcal{C}(\mathcal{A})$.

**Theorem 2.1. (Alexander and Pyke** [1]**)** *If $((\mathbf{X}_n)_{n\geq 1}$ is a sequence of random matrices taking values in $\mathbb{R}^{n\times n}$, whose components are independent and identically distributed random variables having mean zero and variance $\sigma^2$, $0 < \sigma^2 < \infty$, then*

$$\frac{1}{\sigma}\mathbf{T}_n(\mathbf{X}_{n\times n}) \xrightarrow{\mathcal{D}} Z,\ in\ \mathcal{C}(\mathcal{A}),\ as\ n \to \infty,$$

*where $Z := \{Z(A) :\ A \in \mathcal{A}\}$ is the set-indexed Brownian sheet, and "$\xrightarrow{\mathcal{D}}$" stands for the convergence in distribution or weak convergence, see Billingsley* [6].

By extending the result given in Berlinet and Agnan [5] and Lifshits [11] into the set-indexed Brownian sheet, the corresponding RKHS of $Z$ is

$$\mathcal{H}_Z := \left\{h : \mathcal{A} \to \mathbb{R} |\ \exists f \in L_2(\lambda_I) \text{ such that } h(A) = \int_A f\ d\lambda_I,\ A \in \mathcal{A}\right\},$$

where $L_2(\lambda_I)$ is the space of real-valued functions which are squared integrable on $I$ with respect to $\lambda_I$. In the sequel $h_f$ will always denote an absolutely continuous set function on $\mathcal{A}$ with an $L_2(\lambda_I)$ density $f$. The space $\mathcal{H}_Z$ is also a Hilbert space with respect to the inner product and the corresponding norm defined by

$$\langle h_{f_1}, h_{f_2}\rangle_{\mathcal{H}_Z} :=\langle f_1, f_2\rangle_{L_2} = \int_I f_1 f_2 \; d\lambda_I, \;\; \forall h_{f_i} \in \mathcal{H}_Z, i = 1, 2,$$

$$\|h_f\|^2_{\mathcal{H}_Z} := \|f\|^2_{L_2} := \int_I |f|^2 \;\; d\lambda_I.$$

Let $n \geq 1$ be fixed. For every $\mathbf{A}_n, \mathbf{B}_n \in \mathbb{R}^{n\times n}$, it holds $\mathbf{T}_n(\mathbf{A}_n) \in \mathcal{H}_Z$, and

$$\langle \mathbf{A}_n, \mathbf{B}_n\rangle_{\mathbb{R}^{n\times n}} = \langle \mathbf{T}_n(\mathbf{A}_n), \mathbf{T}_n(\mathbf{B}_n)\rangle_{\mathcal{H}_Z}. \tag{2}$$

Eq. (2) can also be verified analogously as in the results in [7, 8]. In particular for large enough $n$, $\mathbf{W}_n$ is isometric with $\mathbf{W}_{n\mathcal{H}_Z} := \mathbf{T}_n(\mathbf{W}_n)$, where $\mathbf{T}_n(\mathbf{W}_n) := [\mathbf{T}_n(f_1(\Xi_n)), \ldots, \mathbf{T}_n(f_p(\Xi_n))] \subset \mathcal{H}_Z$. Hence by Eq. (1) and Eq. (2),

$$\mathbf{T}_n(\mathbf{R}_n) = \mathbf{T}_n(\mathbf{P}_{\mathbf{W}_n^\perp}\mathbf{E}_n) = \mathbf{T}_n(\mathbf{E}_n) - \mathbf{P}_{\mathbf{W}_{n\mathcal{H}_Z}}\mathbf{T}_n(\mathbf{E}_n). \tag{3}$$

The basis of $\mathbf{W}_{n\mathcal{H}_Z}$ can also be constructed in the following manner. For $i = 1, \ldots, p$, let $s_i^{(n)} := \sum_{k=1}^n \sum_{\ell=1}^n f_i(\ell/n, k/n)\mathbf{1}_{C_{\ell k}}$ be a step function on $I$, where $\mathbf{1}_{C_{\ell k}}$ is the indicator of the half-open rectangle $C_{\ell k}$. Then by the preceding result the set of absolutely continuous set functions $\{h_{s_1^{(n)}}, \ldots, h_{s_p^{(n)}}\}$ becomes a basis of $\mathbf{W}_{n\mathcal{H}_Z}$, because it satisfies the condition $h_{s_i^{(n)}} = \frac{1}{n}\mathbf{T}_n(f_i(\Xi_n))$. Without loss of generality we assume in the following that $\{\tilde{f}_1, \ldots, \tilde{f}_p\}$ and $\{\tilde{s}_1^{(n)}, \ldots, \tilde{s}_p^{(n)}\}$ are the Gram-Schmidt orthonormal bases (ONB) of $\mathbf{W}$ and a subspace in $L_2(\lambda_I)$, respectively, associated with $\{f_1, \ldots, f_p\}$ and $\{s_1^{(n)}, \ldots, s_p^{(n)}\}$, respectively. Then by the definition $\{h_{\tilde{f}_1}, \ldots, h_{\tilde{f}_p}\}$ and $\{h_{\tilde{s}_1^{(n)}}, \ldots, h_{\tilde{s}_p^{(n)}}\}$ are the associated Gram-Schmidt ONB of $\mathbf{W}_{\mathcal{H}_Z}$ and $\mathbf{W}_{n\mathcal{H}_Z}$, respectively. Hence the orthogonal projection of any set function $u \in \mathcal{H}_Z$ to $\mathbf{W}_{\mathcal{H}_Z}$ and $\mathbf{W}_{n_{\mathcal{H}_Z}}$, respectively with respect to these bases are represented by $\mathbf{P}_{\mathbf{W}_{\mathcal{H}_Z}}u = \sum_{i=1}^p \langle h_{\tilde{f}_i}, u\rangle_{\mathcal{H}_Z} h_{\tilde{f}_i}$ and $\mathbf{P}_{\mathbf{W}_{n_{\mathcal{H}_Z}}}u = \sum_{i=1}^p \langle h_{\tilde{s}_i^{(n)}}, u\rangle_{\mathcal{H}_Z} h_{\tilde{s}_i^{(n)}}$, respectively. Furthermore, if $f_i$ is continuous on $I$, then $\left\|h_{\tilde{s}_i^{(n)}} - h_{\tilde{f}_i}\right\|_{\mathcal{H}_Z}$ and $\left\|h_{\tilde{s}_i^{(n)}} - h_{\tilde{f}_i}\right\|_{\mathcal{A}}$ converge to zero, as $n \to \infty$, by the fact $\left\|\tilde{s}_i^{(n)} - \tilde{f}_i\right\|_\infty$ converges to zero, as $n \to \infty$. A direct consequence of this results is that: if $(u_n)_{n\geq 1}$ converges to $u$ in $\mathcal{H}_Z$, i.e. converges with respect to $\|\cdot\|_{\mathcal{H}_Z}$, then $\left\|\mathbf{P}_{\mathbf{W}_{n_{\mathcal{H}_Z}}}u_n - \mathbf{P}_{\mathbf{W}_{\mathcal{H}_Z}}u\right\|_{\mathcal{A}}$ converges to zero, as $n \to \infty$.

To be able to project $Z$, we extend the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}_Z}$ into a mapping $\langle\cdot,\cdot\rangle$ defined on the space $\mathcal{H}_Z \times \mathcal{C}(\mathcal{A})$, such that for every $(h_f, u) \in \mathcal{H}_Z \times \mathcal{C}(\mathcal{A})$, $\langle h_f, u\rangle := \int_I^{(C)} f \; du$. The integral $\int_I^{(C)} f \; du$ is the Choquet integral of $f$ with respect to $u$ over $I$, defined by

$$\int_B^{(C)} f \; du := \int_0^\infty u(B\cap\{f \geq x\}) \; dx + \int_{-\infty}^0 [u(B\cap\{f \geq x\}) - u(B)] \; dx, \; B \in \mathcal{A}, \tag{4}$$

where $\{f \geq x\}$ stands for the set $\{\omega \in I : f(\omega) \geq x\}$. The integral on the right-hand side of Eq. (4) is the usual Lebesgue integral. By the definition and by the integration by parts, in case where $u$ is a signed measure on $\mathcal{A}$, the Choquet integral coincides with the Lebesgue integral [14, 15, 19]. Hence, if $u \in \mathcal{H}_Z$ with the $L_2(\lambda_I)$ density $\hat{u}$, we get

$$\langle h_f, u\rangle = \int_I^{(C)} f\ du = \int_I f\ du = \int_I f\hat{u}\ d\lambda_I = \langle h_f, u\rangle_{\mathcal{H}_Z}.$$

This means that $\mathbf{P}_{\mathbf{W}_{n\mathcal{H}_Z}}$ as well as $\mathbf{P}_{\mathbf{W}_{\mathcal{H}_Z}}$ are the restrictions of the set functions $\mathbf{P}^*_{\mathbf{W}_{n\mathcal{H}_Z}} : \mathcal{C}(\mathcal{A}) \to \mathbb{R}$ and $\mathbf{P}^*_{\mathbf{W}_{\mathcal{H}_Z}} : \mathcal{C}(\mathcal{A}) \to \mathbb{R}$, to $\mathcal{H}_Z$, respectively, where for $u \in \mathcal{C}(\mathcal{A})$, $\mathbf{P}^*_{\mathbf{W}_{n\mathcal{H}_Z}} u = \sum_{i=1}^p \langle h_{\tilde{s}_i^{(n)}}, u\rangle h_{\tilde{s}_i^{(n)}}$ and $\mathbf{P}^*_{\mathbf{W}_{\mathcal{H}_Z}} u = \sum_{i=1}^p \langle h_{\tilde{f}_i}, u\rangle h_{\tilde{f}_i}$. The mapping $\mathbf{P}^*_{\mathbf{W}_{\mathcal{H}_Z}}$ is a continuous projections onto $\mathbf{W}_{\mathcal{H}_Z}$ in the sense of Definition 5.15 in [16], since they are linear, continuous, idempotent and surjective.

**Theorem 2.2.** *Suppose that the regression functions $f_1, \dots, f_p$ are continuous and have bounded variation in the sense of Vitali* [9] *on $I$, then under $H_0$ it holds*

$$\frac{1}{\sigma}\mathbf{T}_n(\mathbf{R}_n) \xrightarrow{\mathcal{D}} Z_{\tilde{\mathbf{f}}} := Z - \mathbf{P}^*_{\mathbf{W}_{\mathcal{H}_Z}} Z, \ \text{in } \mathcal{C}(\mathcal{A}), \ \text{as } n \to \infty,$$

*where for every $B \in \mathcal{A}$, $Z_{\tilde{\mathbf{f}}}(B) = Z(B) - \sum_{i=1}^p \langle h_{\tilde{f}_i}, Z\rangle h_{\tilde{f}_i}(B)$.*

*Proof.* Let $(u_n)_{n\geq 1}$ be a sequence in $\mathcal{C}(\mathcal{A})$ such that $\|u_n - u\|_{\mathcal{A}} \to 0$, as $n \to \infty$. By the three angles inequality and by the fact $\left\|h_{\tilde{s}_i^{(n)}}\right\|_{\mathcal{A}} \leq M_i$, for some positive, real number $M_i$, $i = 1, \dots, p$, we get

$$\begin{aligned}
\left\|\mathbf{P}^*_{\mathbf{W}_{n\mathcal{H}_Z}} u_n - \mathbf{P}_{\mathbf{W}^*_{\mathcal{H}_Z}} u\right\|_{\mathcal{A}} &= \left\|\sum_{i=1}^p \langle h_{\tilde{s}_i^{(n)}}, u_n\rangle h_{\tilde{f}_i^{(n)}} - \sum_{i=1}^p \langle h_{\tilde{f}_i}, u\rangle h_{\tilde{f}_i}\right\|_{\mathcal{A}} \\
&\leq \sum_{i=1}^p \Big( \left|\langle h_{\tilde{s}_i^{(n)}}, u_n - u\rangle\right| \left\|h_{\tilde{s}_i^{(n)}}\right\|_{\mathcal{A}} + \left|\langle h_{\tilde{s}_i^{(n)}}, u\rangle - \langle h_{\tilde{f}_i}, u\rangle\right| \left\|h_{\tilde{s}_i^{(n)}}\right\|_{\mathcal{A}} \\
&\qquad + \left|\langle h_{\tilde{f}_i}, u\rangle\right| \left\|h_{\tilde{s}_i^{(n)}} - h_{\tilde{f}_i}\right\|_{\mathcal{A}}\Big) \\
&\leq \sum_{i=1}^p \Big( \left|\langle h_{\tilde{s}_i^{(n)}}, u_n - u\rangle\right| M_i + \left|\langle h_{\tilde{s}_i^{(n)}}, u\rangle - \langle h_{\tilde{f}_i}, u\rangle\right| M_i \\
&\qquad + \left|\langle h_{\tilde{f}_i}, u\rangle\right| \left\|h_{\tilde{s}_i^{(n)}} - h_{\tilde{f}_i}\right\|_{\infty}\Big). \qquad (5)
\end{aligned}$$

For every $n \geq 1$, $\tilde{s}_i^{(n)}$ is $\mathcal{A}-$ measurable and bounded by some positive, real number $N_i$. Hence, we get

$$\left|\langle h_{\tilde{s}_i^{(n)}}, u_n - u\rangle\right| M_i \leq 3N_i \|u_n - u\|_{\mathcal{A}} M_i \to 0, \ \text{as } n \to \infty.$$

Next we consider the second term on the right-hand side of (5). Since $\tilde{s}_i^{(n)}$ converges uniformly to $\tilde{f}_i$, the corresponding sequences of sets $\{\tilde{s}_i^{(n)} \geq t\}$, clearly converges with respect to the symmetric difference metric $d_{\lambda_I}$ to the set $\{f_i \geq t\}$, for every $t \in [-N_i, N_i]$. Hence by the continuity of $u$ on $\mathcal{A}$, $F_n$ converges to $F$, point-wise, as $n \to \infty$, where $F_n(t) := u(\{\tilde{s}_i^{(n)} \geq t\})$ and $F(t) := u(\{\tilde{f}_i \geq t\})$. Furthermore we have $|F_n(t)| \leq \|u\|_{\mathcal{A}}$ and $|F_n(t) - u(I)| \leq 2\|u\|_{\mathcal{A}}$, for every $n \geq 1$ and every

$t \in [-N_i, N_i]$, by the fact $u$ is bounded by $\|u\|_{\mathcal{A}}$ on $\mathcal{A}$. Hence, the Lebesgue bounded convergence theorem [3] implies

$$\left|\langle h_{\tilde{s}_i^{(n)}}, u\rangle - \langle h_{\tilde{f}_i}, u\rangle\right| M_i \to 0, \text{ as } n \to \infty.$$

For the last term we have

$$\left|\langle h_{\tilde{f}_i}, u\rangle\right| \left\|\tilde{s}_i^{(n)} - \tilde{f}_i\right\|_\infty \leq 3\,\|u\|_{\mathcal{A}}\, N_i \left\|\tilde{s}_i^{(n)} - \tilde{f}_i\right\|_\infty \to 0, \text{ as } n \to \infty.$$

The combination of the preceding convergence result in $\left\|\mathbf{P}^*_{\mathbf{W}_{n\mathcal{H}_Z}} u_n - \mathbf{P}_{\mathbf{W}^*_{\mathcal{H}_Z}} u\right\|_{\mathcal{A}} \to 0$, as $n \to \infty$. The proof of the theorem is complete by Eq. (3), Theorem 2.1 and the mapping theorem of Rubyn in Bilingsley [6]. $\square$

**Theorem 2.3.** *The process $Z_{\tilde{\mathbf{f}}}$ is a centered Gaussian process with the covariance function*

$$\begin{aligned}K_{Z_{\tilde{\mathbf{f}}}}(A,B) &:= Cov(Z_{\tilde{\mathbf{f}}}(A), Z_{\tilde{\mathbf{f}}}(B)),\ A, B \in \mathcal{A}\\ &= \lambda_I(A\cap B) - 2\sum_{i=1}^p h_{\tilde{f}_i}(A)h_{\tilde{f}_i}(B) + \sum_{i=1}^p\sum_{j=1}^p h_{\tilde{f}_i}(A)W_{ij}h_{\tilde{f}_j}(B),\end{aligned}$$

*where for $1 \leq i, j \leq p$,*

$$\begin{aligned}W_{ij} := &\int_0^\infty\int_0^\infty \lambda_I(\{\tilde{f}_i \geq s\}\cap\{\tilde{f}_j \geq t\})\ dtds\\ &+ \int_0^\infty\int_{-\infty}^0 \lambda_I(\{\tilde{f}_i \geq s\}\cap\{\tilde{f}_j \geq t\}^c)\ dtds\\ &+ \int_{-\infty}^0\int_0^\infty \lambda_I(\{\tilde{f}_i \geq s\}^c\cap\{\tilde{f}_j \geq t\})\ dtds\\ &+ \int_{-\infty}^0\int_{-\infty}^0 \lambda_I(\{\tilde{f}_i \geq s\}^c\cap\{\tilde{f}_j \geq t\}^c)\ dtds.\end{aligned}$$

*Proof.* By the bi-linearity of the covariance operator, for $A, B \in \mathcal{A}$, we have

$$\begin{aligned}K_{Z_{\tilde{\mathbf{f}}}}(A,B) &:= Cov(Z_{\tilde{\mathbf{f}}}(A), Z_{\tilde{\mathbf{f}}}(B))\\ &= Cov(Z(A) - \sum_{i=1}^p\langle h_{\tilde{f}_i}, Z\rangle h_{\tilde{f}_i}(A), Z(B) - \sum_{j=1}^p\langle h_{\tilde{f}_j}, Z\rangle h_{\tilde{f}_j}(B))\\ &= Cov(Z(A), Z(B)) - Cov(Z(A), \sum_{j=1}^p\langle h_{\tilde{f}_j}, Z\rangle h_{\tilde{f}_j}(B))\\ &\quad - Cov(\sum_{i=1}^p\langle h_{\tilde{f}_i}, Z\rangle h_{\tilde{f}_i}(A), Z(B))\\ &\quad + Cov(\sum_{i=1}^p\langle h_{\tilde{f}_i}, Z\rangle h_{\tilde{f}_i}(A), \sum_{j=1}^p\langle h_{\tilde{f}_j}, Z\rangle h_{\tilde{f}_j}(B)).\end{aligned}$$

Clearly $Cov(Z(A), Z(B)) = \lambda_I(A \cap B)$. The definition of the Choquet integral gives

$$
\begin{aligned}
&Cov(Z(A), \sum_{j=1}^{p} \langle h_{\tilde{f}_j}, Z\rangle h_{\tilde{f}_j}(B)) \\
=& \sum_{j=1}^{p} Cov(Z(A), \int_0^\infty Z(\{\tilde{f}_j \ge t\})dt - \int_{-\infty}^0 \Big[Z(\{\tilde{f}_j \ge t\}) - Z(I)\Big] dt) h_{\tilde{f}_j}(B) \\
=& \sum_{j=1}^{p} (\int_0^\infty \lambda_I(\{\tilde{f}_j \ge t\} \cap A)dt - \int_{-\infty}^0 \Big[\lambda_I(\{\tilde{f}_j \ge t\} \cap A) - \lambda_I(A)\Big] dt) h_{\tilde{f}_j}(B) \\
=& \sum_{j=1}^{p} (\int_A^{(C)} \tilde{f}_j \; d\lambda_I) h_{\tilde{f}_j}(B) = \sum_{j=1}^{p} (\int_A \tilde{f}_j \; d\lambda_I) h_{\tilde{f}_j}(B) = \sum_{j=1}^{p} h_{\tilde{f}_j}(A) h_{f\tilde{f}_j}(B).
\end{aligned}
$$

Analogously, we get

$$
Cov(\sum_{i=1}^{p} \langle h_{\tilde{f}_i}, Z\rangle h_{\tilde{f}_i}(A), Z(B)) = \sum_{i=1}^{p} h_{\tilde{f}_i}(A) h_{\tilde{f}_i}(B).
$$

Furthermore, since $Z(\{\tilde{f}_i \ge t\}) - Z(I) = Z(\{\tilde{f}_i \ge t\}^c)$, then by the definition it holds:

$$
\begin{aligned}
W_{ij} &:= Cov(\langle h_{\tilde{f}_i}, Z\rangle, \langle h_{\tilde{f}_j}, Z\rangle) \\
&= Cov(\int_0^\infty Z(\{\tilde{f}_i \ge s\})ds, \int_0^\infty Z(\{\tilde{f}_j \ge t\})dt) \\
&\quad + \; Cov(\int_0^\infty Z(\{\tilde{f}_i \ge s\})ds, \int_{-\infty}^0 Z(\{\tilde{f}_j \ge t\}^c)dt) \\
&\quad + \; Cov(\int_{-\infty}^0 Z(\{\tilde{f}_i \ge s\}^c)ds, \int_0^\infty Z(\{\tilde{f}_j \ge t\})dt) \\
&\quad + \; Cov(\int_{-\infty}^0 Z(\{\tilde{f}_i \ge s\}^c)ds, \int_{-\infty}^0 Z(\{\tilde{f}_j \ge t\}^c)dt) \\
&= \int_0^\infty \int_0^\infty \lambda_I(\{\tilde{f}_i \ge s\} \cap \{\tilde{f}_j \ge t\}) \; dtds \\
&\quad + \int_0^\infty \int_{-\infty}^0 \lambda_I(\{\tilde{f}_i \ge s\} \cap \{\tilde{f}_j \ge t\}^c) \; dtds \\
&\quad + \int_{-\infty}^0 \int_0^\infty \lambda_I(\{\tilde{f}_i \ge s\}^c \cap \{\tilde{f}_j \ge t\}) \; dtds \\
&\quad + \int_{-\infty}^0 \int_{-\infty}^0 \lambda_I(\{\tilde{f}_i \ge s\}^c \cap \{\tilde{f}_j \ge t\}^c) \; dtds.
\end{aligned}
$$

Hence,

$$Cov(\sum_{i=1}^{p}\langle h_{\tilde{f}_i}, Z\rangle h_{\tilde{f}_i}(A), \sum_{j=1}^{p}\langle h_{\tilde{f}_j}, Z\rangle h_{\tilde{f}_j}(B)) = \sum_{i=1}^{p}\sum_{i=j}^{p} h_{\tilde{f}_i}(A) W_{ij} h_{\tilde{f}_j}(B).$$

By combining all of these results we get the covariance function of $Z_{\mathbf{f}}$ as stated in the theorem. □

Now let us consider the local alternatives $H_1 : Y_n = g^{loc}(\Xi_n) + \mathbf{E}_n$. The set-index RPSP of the observation under $H_1$ is

$$\mathbf{T}_n(\mathbf{R}_n) = \mathbf{T}_n(g^{loc}(\Xi_n)) - \mathbf{P}_{\mathbf{W}_{n\mathcal{H}_Z}}\mathbf{T}_n(\mathbf{g}^{loc}(\Xi_n)) + \mathbf{T}_n(\mathbf{E}_n) - \mathbf{T}_n(\mathbf{P}_{\mathbf{W}_n}\mathbf{E}_n).$$

If $g$ is a continuous and has bounded variation in the sense of Vitaly on $I$, then $\mathbf{T}_n(g^{loc}(\Xi_n))$ converges to $h_g$ uniformly. Consequently by the analogous argument as in the proof of Theorem 2.2 we get $\mathbf{P}_{\mathbf{W}_{n\mathcal{H}_Z}}\mathbf{T}_n(\mathbf{g}^{loc}_{\Xi_n})$ converges to $\mathbf{P}_{\mathbf{W}_{\mathcal{H}_Z}} h_g$ uniformly. Thus by combining this result and Theorem 2.2 we obtain the following corollary.

**Corollary 2.4.** *Suppose that the regression functions $f_1, \ldots, f_p$ are continuous and have bounded variations in the sense of Vitali on $I$. If $g$ has bounded variation in the sense of Vitali and is continuous on $I$, then under the local alternative $H_1 : Y_n = g^{loc}(\Xi_n) + \mathbf{E}_n$ it holds*

$$\frac{1}{\sigma}\mathbf{T}_n(\mathbf{R}_n) \xrightarrow{\mathcal{D}} \frac{1}{\sigma}\varphi_g + Z_{\tilde{\mathbf{f}}}, \ \text{as } n \to \infty,$$

*where $\varphi_g := h_g - \mathbf{P}_{\mathbf{W}_{H_Z}} h_g \in \mathcal{H}_Z$.*

**Remark 2.5.** *Without loss of generality the variance $\sigma^2$ can be assumed to be known, since otherwise, without altering the asymptotic results, $\sigma^2$ can be replaced by a consistent estimator for $\sigma^2$. As for an example we can consider $\hat{\sigma}^2 := \|\mathbf{R}_n\|^2_{\mathbb{R}^{n\times n}} / (n^2 - p)$ which converges to $\sigma^2$ in probability, see [2]. The result in this paper can be directly extended to the $d-$ dimensional unit cube.*

## 3. Examples

**Example 3.1.** *As a first example let us consider the situation for which the model is a zero model $\mathbf{Y}_n = \mathbf{E}_n$. The regression function involved in this model is only zero function. Hence the limit process of the set-indexed RPSP associated to this model is $Z$.*

**Example 3.2.** *For the second example we consider a constant regression model taking the form $\mathbf{Y}_n = \beta + \mathbf{E}_n$, where $\beta$ is an unknown parameter. The only one regression function involved is $\tilde{f}_1(t, s) = 1$, $(t, s) \in I$, which results $h_{\tilde{f}_1}(A) = \lambda_I(A)$,*

*$A \in \mathcal{A}$, and*

$$\int_I^{(C)} \tilde{f}_1 dZ = \int_0^\infty Z(\{\tilde{f}_1 \geq x\})dx + \int_{-\infty}^0 [Z(\{\tilde{f}_1 \geq x\}) - Z(I)]dx$$
$$= \int_0^1 Z(I)dx + \int_1^\infty Z(\emptyset)dx + \int_{-\infty}^0 [Z(I) - Z(I)]dx$$
$$= \int_0^1 Z(I)dx = Z(I),$$

*by the fact $Z(\emptyset) = 0$, almost surely. Hence, the associated limit process is given by $Z_{\tilde{\mathbf{f}}_1}(A) := Z(A) - Z(I)\lambda_I(A)$, for $A \in \mathcal{A}$, which is the set-indexed Brownian bridge, see also* [18].

**Example 3.3.** *Let us consider a first-order regression model as for the third example which takes the form $\mathbf{Y}_n = \sum_{i=1}^3 \beta_i f_i(\Xi_n) + \mathbf{E}_n$, where $\beta_1, \beta_2$, and $\beta_3$ are unknown parameters, and for $(t,s) \in I$, $f_1(t,s) = 1$, $f_2(t,s) = t$ and $f_3(t,s) = s$. The Gram-Schmidt ONB of $\mathbf{W}$ associated to this model is $\tilde{f}_1(t,s) = 1$, $\tilde{f}_2(t,s) = \sqrt{3}(2t-1)$, and $\tilde{f}_3(t,s) = \sqrt{3}(2s-1)$, which satisfy $-\sqrt{3} \leq \tilde{f}_2, \tilde{f}_3 \leq \sqrt{3}$. By the last Example we need only to calculate the Choquet integral of $\tilde{f}_2$ and $\tilde{f}_3$ with respect to the sample path of $Z$. By the definition we have*

$$\int_I^{(C)} \tilde{f}_2 dZ = \int_0^{\sqrt{3}} Z(\{\tilde{f}_2 \geq x\})dx + \int_{\sqrt{3}}^\infty Z(\{\tilde{f}_2 \geq x\})dx$$
$$+ \int_{-\sqrt{3}}^0 [Z(\{\tilde{f}_2 \geq x\}) - Z(I)]dx + \int_{-\infty}^{-\sqrt{3}} [Z(\{\tilde{f}_2 \geq x\}) - Z(I)]dx$$
$$= \int_0^{\sqrt{3}} Z(\{\tilde{f}_2 \geq x\})dx + \int_{\sqrt{3}}^\infty Z(\emptyset)dx$$
$$+ \int_{-\sqrt{3}}^0 [Z(\{\tilde{f}_2 \geq x\}) - Z(I)]dx + \int_{-\infty}^{-\sqrt{3}} [Z(I) - Z(I)]dx$$
$$= \int_0^{\sqrt{3}} Z(\{\tilde{f}_2 \geq x\})dx + \int_{-\sqrt{3}}^0 [Z(\{\tilde{f}_2 \geq x\}) - Z(I)]dx.$$

*Analogously, $\int_I^{(C)} \tilde{f}_3 dZ = \int_0^{\sqrt{3}} Z(\{\tilde{f}_3 \geq x\})dx + \int_{-\sqrt{3}}^0 [Z(\{\tilde{f}_3 \geq x\}) - Z(I)]dx$. Hence, the limit process is presented by*

$$Z_{\tilde{\mathbf{f}}_2}(A) = Z_{\tilde{\mathbf{f}}_1}(A) - \left( \int_0^{\sqrt{3}} Z(\{\tilde{f}_2 \geq x\})dx + \int_{-\sqrt{3}}^0 [Z(\{\tilde{f}_2 \geq x\}) - Z(I)]dx \right) h_{\tilde{f}_2}(A)$$
$$- \left( \int_0^{\sqrt{3}} Z(\{\tilde{f}_3 \geq x\})dx + \int_{-\sqrt{3}}^0 [Z(\{\tilde{f}_3 \geq x\}) - Z(I)]dx \right) h_{\tilde{f}_3}(A).$$

**Example 3.4.** *A Full second-order regression model is a model that has the form $\mathbf{Y}_n = \sum_{i=1}^6 \beta_i f_i(\Xi_n) + \mathbf{E}_n$, where for $(t,s) \in I$, $f_1(t,s) = 1$, $f_2(t,s) = t$, $f_3(t,s) = s$, $f_4(t,s) = t^2$, $f_5(t,s) = s^2$, and $f_6(t,s) = ts$. The Gram-Schmidt ONB of $\mathbf{W}$ is* *then given by* $\tilde{f}_1(t,s) = 1$, $\tilde{f}_2(t,s) = \sqrt{3}(2t-1)$, $\tilde{f}_3(t,s) = \sqrt{3}(2s-1)$, $\tilde{f}_4(t,s) = \sqrt{5}(6t^2 - 6t + 1)$, $\tilde{f}_5(t,s) = \sqrt{5}(6s^2 - 6s + 1)$, *and* $\tilde{f}_6(t,s) = \frac{1}{3}(4ts - 2t - 2s + 1)$. *Since the limit process of the model that only involves* $\tilde{f}_1$, $\tilde{f}_2$, $\tilde{f}_3$ *is already calculated in Example 3.3, we just need in the present one to calculate the Choquet integral of* $\tilde{f}_4$, $\tilde{f}_5$ *and* $\tilde{f}_6$. *We have* $-\sqrt{5}/2 \leq \tilde{f}_4, \tilde{f}_5 \leq \sqrt{5}$ *and* $-1/3 \leq \tilde{f}_6 \leq 1/3$. *Hence, by the definition of Choquet integral, we get*

$$\int_I^{(C)} \tilde{f}_4 dZ = \int_0^{\sqrt{5}} Z(\{\tilde{f}_4 \geq x\})dx + \int_{-\sqrt{5}/2}^{0} [Z(\{\tilde{f}_4 \geq x\}) - Z(I)]dx,$$
$$\int_I^{(C)} \tilde{f}_5 dZ = \int_0^{\sqrt{5}} Z(\{\tilde{f}_5 \geq x\})dx + \int_{-\sqrt{5}/2}^{0} [Z(\{\tilde{f}_5 \geq x\}) - Z(I)]dx$$
$$\int_I^{(C)} \tilde{f}_6 dZ = \int_0^{1/3} Z(\{\tilde{f}_6 \geq x\})dx + \int_{-1/3}^{0} [Z(\{\tilde{f}_6 \geq x\}) - Z(I)]dx$$

*Consequently, the set-indexed residual partial sums limit process of this model is then given by*

$$\begin{aligned} Z_{\tilde{\mathbf{f}}_3}(A) := Z_{\tilde{\mathbf{f}}_2}(A) & \\ & - \left( \int_0^{\sqrt{5}} Z(\{\tilde{f}_4 \geq x\})dx + \int_{-\sqrt{5}/2}^{0} [Z(\{\tilde{f}_4 \geq x\}) - Z(I)]dx \right) h_{\tilde{f}_4}(A) \\ & - \left( \int_0^{\sqrt{5}} Z(\{\tilde{f}_5 \geq x\})dx + \int_{-\sqrt{5}/2}^{0} [Z(\{\tilde{f}_5 \geq x\}) - Z(I)]dx \right) h_{\tilde{f}_5}(A) \\ & - \left( \int_0^{1/3} Z(\{\tilde{f}_6 \geq x\})dx + \int_{-1/3}^{0} [Z(\{\tilde{f}_6 \geq x\}) - Z(I)]dx \right) h_{\tilde{f}_6}(A). \end{aligned}$$

**Acknowledgement** The author is grateful to the German Academic Exchange Service (DAAD) for the financial support during the research stay at the Faculty of Mathematics and Geography of Catholic University Eichstaett-Ingolstadt, Germany. The author wishes to thanks Professor Bischoff and colleges for the cooperation.

## References

[1] Alexander, K.S. and Pyke, R., "A uniform central limit theorem for set-indexed partial-sum processes with finite variance", *The Annals of Probability*, **14:2** (1986), 582-597.

[2] Arnold, S.F., *The Theory of Linear Models and Multivariate Analysis*, John Willy & Sons Inc. New-York, 1981.

[3] Athreya, K.B., and Lahiri, S.N.,*Measure Theory and Probability Theory*, Springer Science + Business Media, LLC New York, 2006.

[4] Bass, R.F. and Pyke, R., "The space $D(A)$ and weak convergence for set-indexed processes", *The Annals of Probability*, **13:3** (1985), 860-884.

[5] Berlinet, A. and Agnan, c.T., *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Springer Verlag New York, 2003.

[6] Billingsley, P., *Convergence of Probability Measures* (second edition), Willy & Sons New-York, 1968.

[7] Bischoff, W., "The structure of residual partial sums limit processes of linear regression models", *Theory of Stochastic Processes*, **2:24** (2002), 23-28.

[8] Bischoff, W. and Somayasa, W., "The limit of partial sums process of spatial least squares residuals", *Journal of Multivariate Analysis*, **100** (2009), 2167-2177.

[9] Clarkson, J.A. and Adams, C.R., "On definition of bounded variation for functions of two variables", *Transactions of the American Mathematical Society*, **35:4** (1933), 824-854.

[10] Dudley, R.M., "Sample functions of Gaussian process", *Ann. Probab*, **1** (1973), 66-103.

[11] Lifshits, M.A., *Gaussian Random Function*, Kluwer Academic Publishers, Dordrecht, 1996.

[12] MacNeill, I.B., "Limit processes for sequences of partial sums of regression residuals", *Ann. Probab.*, **6** (1978), 695-698.

[13] MacNeill, I.B. and Jandhyala, V.K., "Change-point methods for spatial data", *Multivariate Enviromental Statistics* eds. by G.P. Patil and C.R. Rao, Elsevier Science Publisher B.V. (1993), 298-306.

[14] Murofushi, T. and Sugeno, M., "A theory of fuzzy measures: representation, the Choquet integral, and null sets", *Journal of Mathematical Analysis and Application*, **159** (1991), 532-549.

[15] Murofushi, T. and Sugeno, M., "Some quantities represented by the Choquet integral", *Fuzzy Sets and Systems*, **56** (1993), 229-235.

[16] Rudin, W., *Functional Analysis* (second edition), McGraw-Hill, Inc. New York, 1991.

[17] Wang, Z., Klir, G.J., and Wang, W., "Monotone set functions defined by Choquet integral", *Fuzzy Sets and System*, **81** (1996), 241-250.

[18] Xie, L. and MacNeill, I.B., "Spatial residual processes and boundary detection", *South African Statist. J.*, **40:1** (2006), 33-53.

[19] Yan, N., Wang, Z. and Chen, Z., "Classification with Choquet integral with respect to signed non-additive measure", *Proceeding of Sevent IEEE International Conference on Data Mining Workshop* (2007), 283-288.