DSSR

Discussion Paper No. 124

Convolutional regression for big spatial data

Yasumasa Matsuda and Xin Yuan

February, 2022

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Convolutional regression for big spatial data

Yasumasa Matsuda and Xin Yuan*

Abstract

Recently it is common to collect big spatial data on a national or continental scale at discrete time points. This paper aims at a regression model when both dependent and independent variables are big spatial data. Regarding spatial data as functions over a region, we propose a functional regression by a parametric convolution kernel together with the least squares estimation on the frequency domain by applying Fourier transform. It can handle massive datasets with asymptotic validations under the mixed asymptotics. The regression is applied to Covid-19 weekly new cases and human mobility collected in city levels all over Japan to find that an increase of human mobility is followed by an increase of Covid-19 new cases in time lag of two weeks.

keywords: CARMA kernels, Convolution, Fourier transform, Irregularly spaced data, Least squares estimation, NTT DoCoMo spatial statistics, Periodogram, Spurious regression.

1 Introduction

With rapid technological innovations, data structures become increasingly complex and big data analysis becomes especially important. Recently, it is common to collect big spatial data on a national or continental scale at discrete time points. For example, NTT DoCoMo, Inc., a Japanese mobile phone company, provides big spatial dataset recording number of people staying at 500 meter meshes all over Japan every one hour, resulting in terabytes scale spatial data.

Regarding a big spatial data recorded at a time point as one sample of dependent or independent variable in regression analysis, we aim at a regression model in this paper that identifies a non-trivial relationship among big spatial data. For example, it is an interesting problem how much increase of human mobility promote Covid-19 infections, which will be carefully examined in the later section.

*Graduate School of Economics and Management, Tohoku University

Let $y_t(s)$ and $x_{t1}(s), \ldots, x_{tp}(s), t = 1, \ldots, T$, be discrete time observations of big spatial data for $s \in S \subset \mathbb{R}^2$, a large scale region such as nation or continent. Regarding them at each time point as a functional data, we consider a regression model,

$$y_t(s) = \Phi_1(x_{t1})(s) + \cdots + \Phi_p(x_{tp})(s) + \varepsilon_t(s), t = 1, \ldots, T,$$

where Φ_j is not a constant in usual regression but a linear operator on $L^2(S)$. This paper specifies it with a convolution, which is given by

$$\Phi_j(x_t)(s) = \int_S \phi_j(s-u)x_t(u)du,$$

where $\phi_j(\cdot)$ is a function on S called a convolution kernel. We call a regression for functional data with a convolution kernel as convolutional regression. Convolution kernel is a special case of Hilbert Schmidt kernel defined by

$$\Psi_j(x_t)(s) = \int_S \psi_j(s,u)x_t(u)du,$$

with the relation of $\phi(s-u) = \psi(s,u)$. A convolution kernel at s puts more weights on closer points to s. Convolutional regression exploits this relation.

Existing studies in the fields of regression for functional data are found in the literatures of functional data analysis. Ramsay and Silverman (2005) introduces functional regression together with the estimation approach of functional principal component analysis (fPCA). It is regarded as a nonparametric approach to identify Hilbert Schmidt kernels with empirically orthogonal functions. There have been plenty of studies on functional regression under varieties of settings. See, for functional regression, Yao et al. (2005), Horvath and Kokoszka (2012), Crambes and Mas (2013), Benatia et al. (2017) and so on, while see, for functional auto-regression, Bosq (2000). Liu et al. (2016) proposes the unique approach for functional auto-regression with a convolutional kernel by a sieve estimation, which is another nonparametric method different from fPCA. Baseline of them supposes a regression or auto-regression with Hilbert-Schmidt kernels, and identifies the kernels by fPCA under the condition of stationarity across time without restrictions of stationarity across space. They basically suppose an application to functional data on $[a, b]$, a fixed compact interval on $\mathbb{R}$, although it is claimed that they can be extended to $S \subset \mathbb{R}^2$.

This paper tries an approach for spatial regression on huge scales as an alternative to existing ones on fPCA from the two viewpoints. One is that our approach employs parametric identification for convolution kernels rather than nonparametric one for Hilbert Schmidt kernels. The other one is that Fourier transform works to estimate convolution kernels in our approach, while empirically orthogonal functions by fPCA does to identify Hilbert Schmidt kernels in existing approaches. It means that our approach requires no singular value decompositions but Fourier basis functions instead. Fourier transform restricts the applications to spatial data that must be stationary across space but need not be across time, while fPCR restricts spatial data that must be stationary across time but need not be across space. As a result, our approach can handle spatio-temporal data nonstationary across time from short to large time periods including pure spatial data without temporal points. Besides our approach does not suffer spurious regression problems indicated by Granger and Newbold (1974), which will be demonstrated in our simulation studies.

One feature of our approach to be stated finally is the asymptotic validation of our estimation procedures on the frequency domain by Fourier transform, which is obtained by extending the frequency domain techniques developed by time series researchers such as Brockwell and Davis (1991, chap. 10), Robinson (1995), Hosoya (1997) and so on, from time series to spatial data. Under the asymptotic scheme called the mixed domain aymptotics (Stein, 1999, pp. 62), the consistency and asymptotic normality are obtained under the condition of stationarity across space. The consistent estimation of the asymptotic covariance matrix, easily evaluated in practice, makes it possible to conduct statistical inference by t tests in the same way with usual regression analysis. It is actually demonstrated in the empirical section that the statistical inference is successfully applied to detect two week lagged effects between human mobility and Covid-19 new cases collected on national scale in Japan.

2 Convolutional regression for spatial data

We introduce a regression model for spatial data on national or continental scales. Convolutional kernels are employed to define a regression between spatial data.

2.1 Parametric convolution kernels

Let $L^2(S)$ be the set of square integrable functions on a region $S \subset \mathbb{R}^2$. We regard spatial data on $S \subset \mathbb{R}^2$ at discrete time of $t = 1, \ldots, T$ as realizations of $L^2(S)$ valued random variables, denoted as $y_t(s)$ and $x_t(s) = (x_{t1}(s), \ldots, x_{tp}(s))'$, dependent and independent variables, respectively.

For $f(s)$, a $L^2(S)$ function, a convolution is an operator that transforms $f(s)$ to $L^2(S)$ function by

$$\int_S \phi(s-u) f(u) du,$$

where $\phi(s)$ is a convolution kernel. A convolution is a special case of Hilbert

Schmidt operator given by

$$\int_S \psi(s,u)f(u)du,$$

where $\psi(s,u)$ is a Hilbert Schmidt kernel.

Applying $\phi_j(s;\theta)$, a parametric convolution kernel, to $x_{tj}(s)$, we define a convolutional regression given by

$$y_t(s) = \sum_{j=1}^{p} \int_S \phi_j(s-u;\theta)x_{tj}(u)du + \varepsilon_t(s), s \in S, \tag{1}$$

where $\varepsilon_t(s)$ is an unobserved error term regarded as $L^2(S)$ valued random variables.

Parametric convolutional kernels are employed in our approach rather than nonparametric Hilbert Schmidt kernels in traditional approaches of functional regression. Traditional approaches mainly target a regression for $L^2[a,b]$, one dimensional functions on a compact region. Although it is claimed that they are extended to that for $L^2(S), Ssubset\mathbb{R}^2$, the extension is not straightforward especially for big spatial data on national or continental scale.

The spatial regression in (1) allows $y_{t-1}(s), \ldots, y_{t-k}(s)$, temporally lagged terms of dependent variable, to be included as independent variables. Specifically, convolutional auto-regression is defined in the same way as

$$y_t(s) = \sum_{j=1}^{k} \int_S \psi_j(s-u;\theta)y_{t-j}(u)du + \sum_{j=1}^{p} \int_S \phi_j(s-u;\theta)x_{tj}(u)du + \varepsilon_t(s).$$

Conditions to confirm stationary across time can be derived by the arguments in Bosq (2000). We omit details here as stationarity across time is not necessary to validate the estimation in Section 3. We include convolutional auto-regression in (1) throughout this paper. Estimation and asymptotic validations in the later sections work for both convolutional regression and auto-regression.

Let us comment on intercept terms in the convolutional regression. We can include a parametric intercept term $c(s;\theta)$ in (1). However, each of independent variables should be adjusted to be zero mean to avoid identification issues indicated by Kokoszka and Reimherr (2017, pp 75).

2.2 Least squares estimation on the frequency domain

This section considers estimation of parameters that describe the parametric convolutional kernels in (1) in practical situations when sampling points are irregularly spaced and may not be identical for each time point. We will propose the least squares estimation on the frequency domain after introducing the one on the spatial domain as the equivalent alternative.

The usual least squares estimation of θ is obtained by minimizing

$$Q_{sp}(\theta) = \sum_{t=1}^{T} \int \left| y_t(s) - \sum_{a=1}^{p} \int \psi_a(s-u;\theta) x_{ta}(u) du \right|^2 ds, \tag{2}$$

which we call is the least squares estimation on the spatial domain.

Let us consider least squares estimation on the frequency domain as an alternative to the one on the spatial domain, both of which are equivalent mathematically. Fourier transform of $f(s) \in L^2(\mathbb{R}^2)$ is defined by

$$\tilde{f}(\omega) = \int_{\mathbb{R}^2} f(s) e^{-is'\omega} ds, \omega \in \mathbb{R}^2.$$

It is well known that Fourier transform of a convolution

$$g(s) = \int_{\mathbb{R}^2} \psi_a(s-u) f(u) du$$

is given by

$$\tilde{g}(\omega) = \tilde{\psi}_a(\omega) \tilde{f}(\omega).$$

Hence, applying the Fourier transform to (1), we obtain

$$\tilde{y}_t(\omega) = \sum_{j=1}^{p} \tilde{\psi}_j(\omega;\theta) \tilde{x}_{tj}(\omega) + \tilde{\varepsilon}_t(\omega), \omega \in \mathbb{R}^2, \tag{3}$$

and we estimate the parameter θ by minimizing

$$Q_{freq}(\theta) = \sum_{t=1}^{T} \int \left| \tilde{y}_t(\omega) - \sum_{j=1}^{p} \tilde{\psi}_j(\omega;\theta) \tilde{x}_{tj}(\omega) \right|^2 d\omega, \tag{4}$$

which we call the least squares estimation on the frequency domain.

Since $Q_{sp}(\theta) = Q_{freq}(\theta)$ by Parseval's theorem, the least squares estimation on both spatial and frequency domains are completely equivalent. Practical situations when we evaluate them by several approximations, however, makes the two least squares estimation be different. This paper will focus the least squares on the frequency domain by two reasons. One is less computational costs of the convolution caused by Fourier transform. The other one is the asymptotic validations in Section 3 that make it possible to conduct statistical inference.

Suppose we observe $y_t(s)$ and $x_{ta}(s)$ on randomly distributed points inside a rectangular $A = [0, A_1] \times [0, A_2]$, where observation points may or may not be identical for $t = 1, \ldots, T$. Let us denote them as $s_{0tj}, j =$

$1, \ldots, n_{0t}$ for $y_t(s)$ and $s_{atj}, j = 1, \ldots, n_{at}$ for $x_{ta}(s)$. Then let us define the discrete Fourier transform (DFT) of $y_t(s)$ and $x_t(s)$ by, for $t = 1, \ldots, T$,

$$\hat{y}_t(\omega) = \frac{1}{n_{0t}} \sum_{j=1}^{n_{0t}} y_t(s_{0tj}) \exp(-i\omega' s_{0tj}),$$

$$\hat{x}_{ta}(\omega) = \frac{1}{n_{at}} \sum_{j=1}^{n_{at}} x_{ta}(s_{atj}) \exp(-i\omega' s_{atj}), a = 1, \ldots, p.$$

Then replacing $\tilde{y}_t(\omega), \tilde{x}_{ta}(\omega)$ in (3) with the DFTs on mesh points of Fourier frequencies,

$$\omega_f = \left(\frac{2\pi f_1}{A_1}, \frac{2\pi f_2}{A_2} \right), (f_1, f_2) \in \mathbb{Z}^2,$$

in a compact region D on $\mathbb{R}^2$, we estimate the parameter θ by minimizing

$$\tilde{Q}_{freq}(\theta) = \sum_{t=1}^{T} \sum_{\omega_f \in D} \left| \hat{y}_t(\omega_f) - \sum_{a=1}^{p} \tilde{\psi}_a(\omega_f; \theta) \hat{x}_{ta}(\omega_f) \right|^2, \tag{5}$$

which we denote as $\hat{\theta}$. $\tilde{Q}_{freq}(\theta)$ is regarded as an empirical evaluation of $Q_{freq}(\theta) = Q_{sp}(\theta)$. We propose the one minimizing $\tilde{Q}_{freq}(\theta)$ as the estimator in the practical situations when sampling points x and y are irregularly spaced and may or may not be identical for each t with possibly time varying sample sizes.

Remark 1. Let us introduce the empirically adjusted version of $Q_{sp}(\theta)$ in (2). Approximating the convolution with Riemannian summation, we estimate θ by minimizing, for observation points $s_{atj}, j = 1, \ldots, N_{at}, a = 0, 1, \ldots, p$,

$$\tilde{Q}_{sp}(\theta) = \sum_{t=1}^{T} \sum_{j=1}^{N_{0t}} \left\{ y_t(s_{0tj}) - \sum_{a=1}^{p} \frac{1}{K_{atj}} \sum_{k=1}^{N_{at}} \psi_a(s_{0tj} - s_{atk}; \theta) x_{at}(s_{atk}) \right\}^2, \tag{6}$$

with

$$K_{atj} = \sum_{k=1}^{N_{at}} \phi_0(s_{0tj} - s_{atk}; \theta).$$

Although $Q_{sp}(\theta) = Q_{freq}(\theta)$ theoretically, their empirical counterparts of $\tilde{Q}_{sp}(\theta)$ and $\tilde{Q}_{freq}(\theta)$ are not equal because of several differences in the approximations. It follows that the two least squares estimation on the spatial and frequency domains are not the same. We claim advantages of the estimation on the frequency domain for big spatial data on huge scales that come from lower computational costs and the asymptotic validations in the following section.

Remark 2. As in usual regression analysis, we often need to construct predicted values. Approximating the convolution in (1) by Riemannian summation, we evaluate the forecast at v and t by

$$\hat{y}_t(v) = \sum_{a=1}^{p} \frac{1}{K_{at}} \sum_{k=1}^{N_{at}} \psi(v - s_{atk}; \hat{\theta}) x_{at}(s_{atk}), \tag{7}$$

$$K_{at} = \sum_{k=1}^{N_{at}} \psi(v - s_{atk}; \hat{\theta}).$$

2.3 Parametric family of CARMA kernels

Suitable parametric family of kernels are necessary to apply the convolutional regression in (1) to real data in practice. They need to satisfy two requirements. First, they cover reasonably broad class of convolutions by a parametric family. Second, they have the closed form expressions of Fourier transforms necessary to estimate the parameters by the least squares on the frequency domain in (5). CARMA kernels, which were proposed in Brockwell and Matsuda (2017) to define random fields as an extension of Ornstein Uhlenbeck processes (Uhlenbeck and Ornstein, 1930), are introduced.

Let $a(z) = \prod_{i=1}^{p}(z^2 - \lambda_i^2)$ with $Re(\lambda_i) < 0$ and $b(z) = \prod_{i=1}^{q}(z^2 - \xi_i^2)$ with $0 \leq q < p$ and $\lambda_i^2 \neq \xi_j^2$ for all i and j. CARMA(p, q) kernel is defined by

$$\phi(s) = \sum_{i=1}^{p} \frac{b(\lambda_i)}{a'(\lambda_i)} \exp(\lambda_i ||s||), s \in \mathbb{R}^2,$$

where a' denotes the derivative of the polynomial a and $||s||$ is the Euclidean norm of he vector s. CARMA kernel is an isotropic function of $||s||$ and has the closed form expressions of Fourier transform given by

$$\tilde{\phi}(\omega) = -\sum_{i=1}^{p} \frac{\lambda_i b(\lambda_i)}{a'(\lambda_i)(||\omega||^2 + \lambda_i^2)^{3/2}}.$$

The simplest kernel of CAR(1), obtained by substituting $p = 1, q = 0$ in the CARMA(p, q) kernel, and the Fourier transform are given by, after transforming λ_1 with $-1/\delta$ for later use,

$$\phi(s; \beta, \delta) = \frac{\beta}{2\pi\delta^2} e^{-||s||/\delta}, \delta > 0, s \in \mathbb{R}^2, \tag{8}$$

$$= \beta \phi_0(s; \delta),\ \ say,$$

$$\tilde{\phi}(\omega; \beta, \delta) = \frac{\beta}{\delta^3(||\omega||^2 + \delta^{-2})^{3/2}}, \omega \in \mathbb{R}^2, \tag{9}$$

$$= \beta \tilde{\phi}_0(\omega; \delta),\ \ say.$$

The CAR(1) kernel here is normalized to let the integral over $\mathbb{R}^2$ be β and make easier interpretation for the parameter β as a result. The convolutional regression with the CAR(1) kernels is expressed as

$$y_t(s) = \sum_{j=1}^{p} \beta_j \int_S \frac{1}{2\pi\delta_j^2} e^{-||s-u||/\delta_j} x_t(u)du + \varepsilon_t(s), \delta > 0, s \in S, \tag{10}$$
$$= \sum_{j=1}^{p} \beta_j \int_S \phi_0(s-u;\delta_j)x_t(u)du + \varepsilon_t(s).$$

The parameter β measures a degree of spatial association between x and y, while δ describes spatial effects that measure how far the neighbours of $x(s)$ affect $y(s)$. As $\delta_j \to 0$, it reduces to

$$y_t(s) = \sum_{j=1}^{p} \beta_j x_{tj}(s) + \varepsilon_t(s), s \in S,$$

the usual regression model without spatial effects. One point to be noticed here is that $\beta = 0$ in (8) makes δ unidentifiable. Testing spatial independence between $x(s)$ and $y(s)$ by $\beta = 0$ should be tried for a fixed value of δ. In the later simulation studies, we tried the test of independence by testing $\beta = 0$ for $\delta = 0$.

3 Asymptotic properties

This section considers asymptotic properties of the least squares estimator in (5) when we have samples of $y_t(s), x_{ta}(s), a = 1, \ldots, p$, on irregularly spaced points on $[0, A_1] \times [0, A_2]$ for $t = 1, \ldots, T$. Three kinds of asymptotics are known in spatial statistics literatures (Stein, 1999). This paper employs the mixed asymptotic among the three when both region of $[0, A_1] \times [0, A_2]$ and number of sampling points diverge jointly in the spatial dimension, while the temporal size of T is finite and fixed in the temporal dimension. The techniques to prove the asymptotic results are based on Matsuda and Yajima (2009) and Matsuda and Yajima (2018), which have developed the method to manage irregularly spaced data under the mixed asymptotics. Let us denote the least squares estimator minimizing (5) as $\hat{\theta}$ throughout this section.

3.1 Consistency and asymptotic normality

Let us clarify the conditions under which consistency and asymptotic normality of our estimator is validated.

C1. The sample sizes n_{at} and the sampling region $A = [0, A_1] \times [0, A_2]$ diverge jointly such that $A_1 \to \infty, A_2 \to \infty, A_1/A_2, A_2/A_1 = O(1)$,

$n_{at}/n_{bt} = 0(1)$ and $|A|/n_{at} \to 0, a, b = 0, 1, \ldots, p, t = 1, \ldots, T$, for the area $|A| = A_1 \times A_2$. We shall employ a suffix k such as $A = A_k, n_{at} = n_{atk}$ when we indicate explicitly that they diverge as k tends to infinity.

C2. Let S_{at} be the set of sampling points in $A = [0, A_1] \times [0, A_2]$. We assume that elements in S_{at} are written as, for $a = 0, 1, \ldots, p, t = 1, \ldots, T$,

$$s_{atj} = (A_1 u_{1,atj}, A_2 u_{2,atj}), j = 1, \ldots, n_{at},$$

where $u_{atj} = (u_{1,atj}, u_{2,atj})'$ is a sequence of independent and identically distributed random vectors with a probability density function $g(x)$ supported on $[0, 1]^2$ which has continuous first derivatives. Sampling points s and observed variables x_t, ε_t at s are independent for all t.

C3. $x_t(s) = (x_{t,1}(s), \ldots, x_{t,p}(s))'$ has finite moments of all orders for any $t = 1, \ldots, T$, $s \in \mathbb{R}^2$, and is stationary across $s \in \mathbb{R}^2$ with all higher order spectral density functions that are all bounded and integrable. The 2nd order spectral density function,

$$f_{x,t_1t_2,a_1a_2}(\omega) = (2\pi)^{-2} \int_{\mathbb{R}^2} Cov(X_{t_1,a_1}(s), X_{t_2,a_2}(s-u))e^{-iu'\omega}du,$$

is positive definite and twice differentiable.

C4. $x_t(s)$ is strictly exogenous in the sense that $x_{t_1}(s_1)$ is independent of $\varepsilon_{t_2}(s_2)$ for any $t_1, t_2 = 1, \ldots, T, s_1, s_2 \in \mathbb{R}^2$. $\varepsilon_t(s)$ has finite moments of all orders, and is stationary across $s \in \mathbb{R}^2$ with all higher order spectral density functions that are all bounded and integrable. The 2nd order spectral density,

$$f_{\varepsilon,t_1t_2}(\omega) = (2\pi)^{-2} \int_{\mathbb{R}^2} Cov(\varepsilon_{t_1}(s), \varepsilon_{t_2}(s-u))e^{-iu'\omega}du,$$

is positive and twice differentiable.

C4'. $x_t(s)$ is contemporaneously exogenous in the sense that $x_t(s_1)$ is independent of $\varepsilon_t(s_2)$ for any $t = 1, \ldots, T, s_1, s_2 \in \mathbb{R}^2$. $\varepsilon_t(s)$ has finite moments of all orders, is serially independent across t and stationary across $s \in \mathbb{R}^2$ with all higher order spectral density functions. The 2nd order spectral density function,

$$f_{\varepsilon,t}(\omega) = (2\pi)^{-2} \int_{\mathbb{R}^2} Cov(\varepsilon_t(s), \varepsilon_t(s-u))e^{-iu'\omega}du,$$

is positive definite and twice differentiable.

C5. Let Θ be a compact parameter space in $\mathbb{R}^q$ for $\theta = (\theta_1, \ldots, \theta_q)$, and D be a fixed symmetric compact region on $\mathbb{R}^2$. $\tilde{\psi}(\omega;\theta)$ has a continuous first derivative with respect to θ on $\Theta \times D$. $\theta_1 \neq \theta_2$ implies that $\tilde{\psi}(\omega;\theta_1) \neq \tilde{\psi}(\omega;\theta_2)$ on a subset of D with positive Lebesgue measure.

C1 clarifies the details of the mixed asymptotics under which the asymptotic properties of the estimator is validated. Notice that the asymptotics in space is employed with fixed and finite temporal size T rather than joint asymptotics of space and time. C2 specifies irregularity of sampling points over the diverging region $A = [0, A_1] \times [0, A_2]$ with the identical density function of $|A|^{-1}g(s/A)$. The distribution of sampling points needs to be independent of dependent and independent variables. It should be emphasized that C3 and C4 (C4') confine $z_t(s) = (y_t(s), x_t(s))$ to be stationary across s without stationary restrictions across t. Namely, $z_t(s)$ can be nonstationary across t. C4 assumes the strict exogeneity of $x_t(s)$ to allow serial correlation of the error terms, while C4' assumes serially independency of the error terms to relax the strict exogeneity with the contemporaneous exogeneity. C5 is the usual assumption of identifiability of parametric convolutional kernels.

Now we are in a position to state the consistency and asymptotic normality. The proofs are shown in the final section of appendix.

Theorem 1. *Under Assumptions C2, C3, C4 and C5,*

$$\hat{\theta} \to_p \theta_0$$

in the asymptotic scheme defined by C1. Assumption C4 can be replaced with C4' to have the consistency.

Theorem 2. *Under Assumptions C2, C3, C4 and C5, if* $|A|^{3/2}/n_{at} = o(1), a = 0, 1, \ldots, p,$

$$\sqrt{|A|}(\hat{\theta} - \theta_0) \to N\left(0, (2\pi)^2 b_g \Omega^{-1} \Sigma \Omega^{-1}\right)$$

in the asymptotic scheme defined by C1, where

$$b_g = \left\{ \int_{[0,1]^2} |g(u)|^4 du \right\} \left\{ \int_{[0,1]^2} |g(u)|^2 du \right\}^{-2},$$

$$\Omega = \int_D \dot{\Psi}(\omega;\theta_0) \left\{ \sum_{t=1}^{T} f_{x,tt}(\omega) \right\} \dot{\Psi}'(\omega;\theta_0) d\omega,$$

$$\Sigma = \int_D \dot{\Psi}(\omega;\theta_0) \left\{ \frac{1}{2} \sum_{t_1,t_2=1}^{T} f_{\varepsilon,t_1t_2}(\omega) f_{x,t_1t_2}(\omega) \right\} \dot{\Psi}'(\omega;\theta_0) d\omega,$$

and where $f_{x,tt}(\omega)$ *is the spectral density matrix whose* (a,b)*th element is* $f_{x,tt,ab}(\omega)$ *in Assumption C3 and* $\dot{\Psi}(\omega;\theta)$ *is the* q *by* p *matrix whose* (i,j)*th* *element is*

$$\frac{\partial \tilde{\psi}_j(\omega;\theta)}{\partial \theta_i}, i = 1, \ldots, q, j = 1, \ldots, p.$$

If Assumption C4' is assigned for Assumption C4, Σ in the asymptotic variance matrix reduces to

$$\Sigma = \int_D \dot{\Psi}(\omega;\theta_0) \left\{ \frac{1}{2} \sum_{t=1}^{T} f_{\varepsilon,tt}(\omega) f_{x,tt}(\omega) \right\} \dot{\Psi}'(\omega;\theta_0) d\omega.$$

We will give a few remarks on the asymptotic results. First, the asymptotic results are validated under the asymptotics when observation region and samples inside together diverge for a fixed time period, which makes it possible to relax stationarity across time under stationarity only across space. Secondly, b_g is the term that clarifies the effects of distributions of sampling points to the asymptotic variance. Sampling points uniformly distributed over a rectangular make the least squares estimation most efficient. Non-uniform distributions over a region that deviates from a rectangular worsen the efficiency. Thirdly, the asymptotic variance corresponds with the heteroskedasticity- robust one in the traditional regression analysis. In other words, the error term in (3) is regarded as the one with heteroskedastic variance on the frequency domain. Finally, $|A|$, the area of the sampling region, is the normalizing factor in the asymptotic distribution. This is the notifying feature in the asymptotics that contrasts with the sample size in the traditional central limit theorems for iid sequences.

3.2 consistent asymptotic variance estimation

Let us consider the consistent estimation of the asymptotic variance matrix in Theorem 2, which shall be used to conduct statistical inference for the parameters in the kernels $\psi_a(u,\theta), a = 1, \ldots, p$. First, we estimate the density $g(x)$ in condition C2 that appears in the asymptotic variance matrix. Let K be a kernel function that is positive and continuous on $[0,1]^2$, and let $K_h(s) = h_1^{-1} h_2^{-1} K(s_1/h, s_2/h), s = (s_1, s_2) \in [0,1]^2$ for $h = (h_1, h_2)$, a bandwdith. Estimate $g(x), x \in [0,1]^2$ by

$$\hat{g}(x) = \frac{1}{(p+1)T} \sum_{a=0}^{p} \sum_{t=1}^{T} \hat{g}_{at}(x),$$

$$\hat{g}_{at}(x) = \frac{1}{n_{at}} \sum_{i=1}^{n_{at}} K_h(s_{ati}/A - x),$$

where $s_{ati}/A = (s_{ati,1}/A_1, s_{ati,2}/A_2)$. Then for $x_i \in [0,1]^2, i = 1, \dots, N^2$, mesh points equally dividing $[0,1]^2$ into N^2 squares, let

$$\hat{b}_g = \frac{\frac{1}{N^2}\sum_{i=1}^{N^2} \hat{g}(x_i)^4}{\left\{\frac{1}{N^2}\sum_{i=1}^{N^2} \hat{g}(x_i)^2\right\}^2}.$$

Then the asymptotic variance matrix evaluated in Theorem 2 as

$$|A|^{-1}(2\pi)^2 b_g \Omega^{-1}\Sigma\Omega^{-1},$$

is consistently estimated by

$$\hat{b}_g \hat{\Omega}^{-1}(\hat{\theta})\hat{\Sigma}(\hat{\theta})\hat{\Omega}(\hat{\theta})^{-1},$$

where

$$\hat{\Omega}(\theta) = \sum_{\omega_f \in D} \dot{\Psi}(\omega_f;\theta)\left\{\sum_{t=1}^{T} \hat{x}_t(\omega_f)\overline{\hat{x}_t(\omega_f)}'\right\}\dot{\Psi}(\omega_f;\theta)', \quad (11)$$

$$\hat{\Sigma}(\theta) = \sum_{\omega_f \in D} \dot{\Psi}(\omega_f;\theta)\left[\sum_{t_1,t_2=1}^{T} \{L_{t_1t_2}(\omega_f;\theta) + M_{t_1t_2}(\omega_f,\theta)\}\right]\dot{\Psi}(\omega_f;\theta)',$$

$$L_{t_1t_2}(\omega_f;\theta) = Re\{\hat{\varepsilon}_{t_1}(\omega_f;\theta)\}Re\{\hat{\varepsilon}_{t_2}(\omega_f;\theta)\}Re\{\hat{x}_{t_1}(\omega_f)\}Re\{\hat{x}_{t_2}(\omega_f)'\},$$

$$M_{t_1t_2}(\omega_f;\theta) = Im\{\hat{\varepsilon}_{t_1}(\omega_f;\theta)\}Im\{\hat{\varepsilon}_{t_2}(\omega_f;\theta)\}Im\{\hat{x}_{t_1}(\omega_f)\}Im\{\hat{x}_{t_2}(\omega_f)'\},$$

$$\hat{\varepsilon}_t(\omega_f;\theta) = \hat{y}_t(\omega_f) - \sum_{a=1}^{p} \tilde{\psi}_a(\omega_f;\theta)\hat{x}_{ta}(\omega_f).$$

The consistent estimator for the asymptotic variance makes it possible to conduct the t test for the least squares estimators. The term of $(2\pi)^2|A|^{-1}$ in the asymptotic variance cancels in the consistent estimation with the Riemannian approximation for the integral.

4 Simulation studies

This section examines empirical performances of the least squares estimation on the frequency domain proposed in Section 2.2. We validated the asymptotic consistency and normality under Assumption C2-C4 in the asymptotic scheme in C1, where C3 assumes stationary only across space, does not across time. Namely, our estimation is validated even for temporally unit root processes. We check the empirical performances under C3 when temporal stationary is not necessarily satisfied.

First, we study the spatial extension of spurious regression problem found by Granger and Newbold (1974). They found that, for y_t and x_t, which are mutually independent unit root processes, the t test for the simple regression analysis rejects the independence more often than significance level, which is the so called spurious regression. Now we shall check it for our least squares estimation.

We simulate $y_t(s), x_t(s)$, which are mutually independent spatio-temporal data on $A = [0, 20]^2$ by spatial moving averages of unit root processes. Let $u_k, v_k, k = 1, \ldots, 2000$ be knots uniformly distributed on $A = [0, 40]^2$ and simulate by

$$y_t(s) = \sum_{k=1}^{2000} \eta_{tk}^{(1)} e^{-||s-u_k||}, \ x_t(s) = \sum_{k=1}^{2000} \eta_{tk}^{(2)} e^{-||s-v_k||}, \tag{12}$$

where $\eta_{t,k}^{(i)}, i = 1, 2$ are mutually independent unit root processes for each k, simulated by

$$\eta_{tk}^{(i)} = \eta_{t-1,k}^{(i)} + \varepsilon_{tk}^{(i)}, t = 1, \ldots, T,$$

where $\varepsilon_{tk}^{(i)}$ are independent iid standard normal variables. We simulated $x_t(s)$ and $y_t(s)$ on two different sets of uniformly distributed 2000 points on A for $t = 1, \ldots, T$. They are mutually independent processes by the simulating models.

For the simulated dataset of $x_t(s)$ and $y_t(s)$ that are mutually independent, we fitted the convolutional regression model in (1) with the CAR(1) kernel in (8). The true kernel function in this case is attained by (8) with $\beta = 0$ for any values of $\delta > 0$, namely δ is not an identifiable parameter in the independent case. Hence we fix δ with 0.5 and 2.0 in the kernel as examples, and estimate β by the least squares in (5) for the fixed $\delta = 0.5, 2.0$, where D is the half circle with the centre of origin on the upper half plane adjusted to make the number of elements to $2000/2 = 1000$. Figure 1 depicts the histograms of the t values of β for $T = 30$ and 200, evaluated by 100 simulations.

Figure 1 demonstrates that the asymptotic theory in section 3.1 works for non-stationary unit root processes across time. Namely, the t values follow distributions well approximated by standard normal one. The sizes of the $t$test are 3, 6, 6 and 6 for the cases of $\delta = 0.5, T = 30$. $\delta = 0.5, T = 200$, $\delta = 2, T = 30$ and $\delta = 2, T = 200$, under the 5% significance level. Hence it is clarified empirically as well as asymptotically that convolutional regressions do not suffer from the spurious regression problems indicated by Granger and Newbold (1974).

Next, let us move on to dependent cases when $y_t(s)$ is given by a convo-

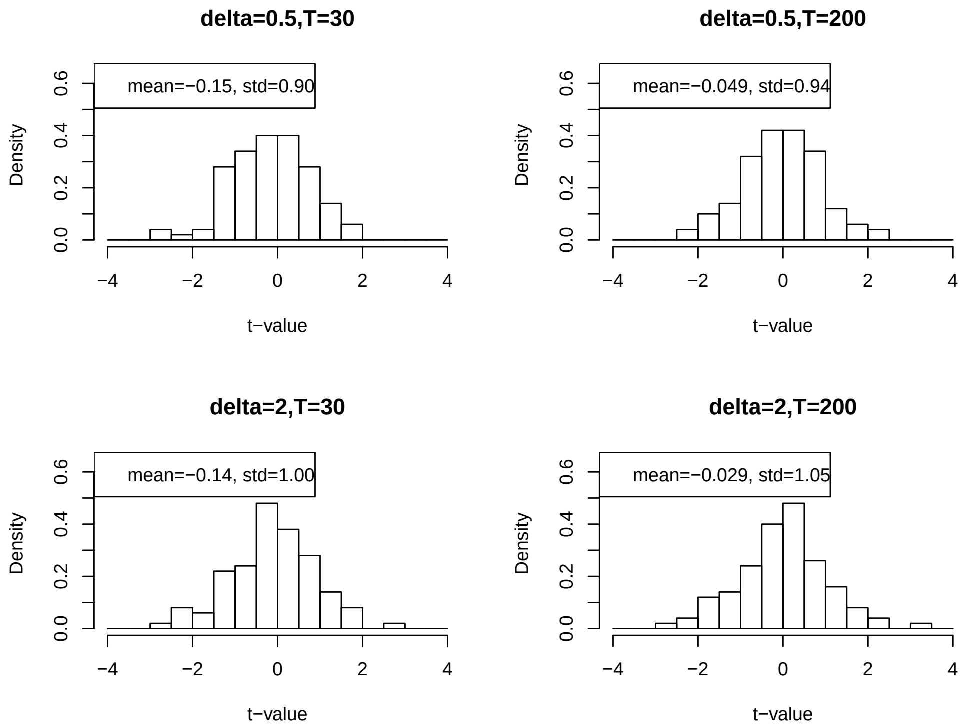

Figure 1: The histograms of t values for the mutually independent spatio-temporal dataset simulated by (12) evaluated by 100 simulations.

lution of $x_t(s)$. We simulated $y_t(s)$ by, for $A = [0, 20]^2$,

$$y_t(s) = \beta \int_A \phi_0(s - u; \delta) x_t(u) du + \varepsilon_t(u), t = 1, \ldots, T,$$

where ϕ_0 is the CAR(1) kernel in (8), and

$$x_t(s) = \sum_{k=1}^{6000} \xi_{tk}^{(1)} e^{-||s-u_k||}, \ \varepsilon_t(s) = \sum_{k=1}^{6000} \xi_{tk}^{(2)} e^{-||s-v_k||},$$
$$\xi_{tk}^{(i)} = \rho \xi_{t-1,k}^{(i)} + f_{tk}^{(i)}, i = 1, 2,$$

and where u_k, v_k are knots uniformly distributed on A and $f_{tk}^{(i)}, i = 1, 2$, are independent standard normal variables. The convolution operator was approximated with the Riemannian summation when we simulated in practice. Notice that the simulated dataset satisfy Assumption C4 for any value of ρ. We have an interest in the estimation performances of β and δ when ρ is designed as 0 and 1 that simulate stationary and unit root across time, respectively.

	$\rho = 0$		$\rho = 1$	
	β	δ	β	δ
mean	1.01	0.49	1.04	0.51
RMSE	0.041	0.023	0.18	0.10
ave(se)	0.061	0.052	0.16	0.11

Table 1: The means and RMSEs of LSE minimizing (5) for $\beta = 1, \delta = 0.5$ in the CAR(1) kernel in the cases of $\rho = 0$ and 1, which simulate stationary and unit root across time, respectively, evaluated by 100 simulations. ave(se) is the average of estimated standard errors in (11).

For 100 sets of $y_t(s), x_t(s)$ simulated for $\beta = 1, \delta = 0.5$ on 6000 uniformly scattered points on A with $T = 30$ in cases 1 and 2 of $\rho = 0$ and 1, respectively. We estimated β, δ in the CAR(1) kernel by minimizing (5). Table 1 lists the average and root mean squared error of $\hat{\beta}$ and $\hat{\delta}$.

Table 1 demonstrates that the least squares by minimizing (5) works efficiently for both cases of $\rho = 0$, error term is stationary across time, and $\rho = 0$, unit root across time. They have small biases and root mean squares that are well approximated by the standard errors in (11) based on the asymptotic variance in Theorem 2. The simulation results validate that the asymptotic distribution approximates the empirical distribution accurately for both cases of stationary and unit root non-stationary of error terms.

5 Real data analysis

In this section, we demonstrate the applications of the convolutional regression model in (1) to two real spatio-temporal data. They are monthly precipitation recorded at US weather stations and weekly number of new cases of COVID-19 in every city all over Japan. Observation points in both examples are irregularly spaced in the way we supposed for the estimation in section 2.2.

Here let us denote several remarks for the demonstrations. First, we shall employ CAR(1) kernels in (8) for all the convolutional regression models in the section. Secondly, all the spatial data we considered for dependent and independent variables at all temporal points were detrended in order to avoid the intercept issues in Section 2.1 as well as as well as to satisfy stationary conditions in Assumption C4 in Section 3. Specifically we detrended separately for each temporal points by fitting CAR(1) random field by Brockwell and Matsuda (2017),

$$\psi_j(s; \alpha) = e^{-\alpha||s-u_j||}, \alpha > 0, s \in \mathbb{R}^2, \tag{13}$$

where $u_j, j = 1, \dots, N$, are knots scattered randomly over the observation region. Finally, we took two benchmarks for comparisons with the forecast-

Figure 2: Weather stations in United States

ing performances. For our forecast with the least squares estimation on the frequency domain in (5), the first benchmark is the predicted value with the estimation on the spatial domain in (6), and the other one is the trend that we identified for the removal of the trend function by CAR(1) random field.

5.1 US weather data

Here we apply the functional regression model in (1) to US precipitation data, which is monthly total precipitation, maximum and minimum temperatures observed at weather stations all over the US from 1895 through 1997, available on the Institute for Mathematics Applied Geosciences (IMAG) website. All the observations are recorded monthly as spatio-temporal data with their longitudes and latitudes. For more details, visit `http://www.image.ucar.edu/Data/US.monthly.met/USmonthlyMet.shtml`. Around 6,000 weather stations are shown in Figure 2.

Regarding the monthly precipitation, and maximum and minimum temperatures as dependent and independent variables, respectively, we fitted the functional spatial regression model in (1) given by, for the CAR(1) kernel in (10),

$$ppt_t(s) = \beta_1 \int \phi_0(s-u;\delta_1) tmin_t(u) du + \beta_2 \int \phi_0(s-u;\delta_2) tmax_t(u) du + \varepsilon_t(s). \tag{14}$$

We took the observations from 1986 to 1990 as in-samples and those in

variable	param.	LSE by (5)	se
min temp.	β_1	1.97	(0.15)
	δ_1	0.62	(0.053)
max temp.	β_2	-1.64	(0.10)
	δ_2	0.38	(0.023)

Table 2: The estimated parameters for (14) with their standard errors, when we took $A = [0, 50] \times [0, 30]$ to construct the periodogram on the mesh points $(2\pi i/A_1, 2\pi j/A_2), i, j \neq 0$ in $D = \{\omega \in \mathbb{R}^2, ||\omega|| < K\}$, where K was adjusted to make the number of the elements be 4000.

	forecasting MSE	
model	in-sample	out-of-sample
CAR(1) by (5)	23.7	28.8
CAR(1) by (6)	23.4	28.6
trend fit by (13)	28.7	36.3

Table 3: Forecasting MSE for in-samples (60 months in 1986-1990) and out-of-samples (12 months in 1991) in US precipitation over 6000 weather stations.

1991 as out-of-samples. The precipitation, maximum, and minimum temperature do not share the same observation points that differ from month to month. Thus the numbers of the observations depend on month. They are around 6000 for precipitation and 4000 for maximum and minimum temperature on average. The total period is 72 months, 60 months for in-samples and 12 months for out-of-samples.

We transformed the coordinate of latitudes and longitudes to record the observation locations into Cartesian coordinate based on global distance with an unit of 100 km. Applying the discrete Fourier transform to the in-sample observations, we estimated the parameters, $\beta_1, \beta_2, \delta_1, \delta_2$ by minimizing the equation in (5) with (9), where we took $A = [0, 50] \times [0, 30]$ to construct the periodogram on the mesh points $(2\pi i/A_1, 2\pi j/A_2), i, j \neq 0$ in $D = \{\omega \in \mathbb{R}^2, ||\omega|| < K\}$, where K was adjusted to make the number of the elements be 4000. The estimation results are shown in Table 2.

Then we constructed the forecasts by (14) in (7) with the estimated parameters to compare the mean squared errors with those of the benchmarks. The MSEs evaluated by, for $v_{kt}, k = 1, \ldots, n_t$, the out-of-samples at $t = 61, \ldots, 72$,

$$\frac{1}{n_{61} + \cdots + n_{72}} \sum_{t=61}^{72} \sum_{k=1}^{n_t} \left(ppt_t(v_{kt}) - \hat{ppt}_t(v_{kt})\right)^2$$

are shown in Table 3.

We find from Table 2 that we identified the non-trivial relations. Namely, precipitation is negatively correlated with maximum temperatures, while it is positively correlated with minimum temperatures. The evaluated standard errors for the estimators indicate that they are strongly significant with the p-values less than 1%. Table 3 demonstrates that our estimation on the frequency domain works as well as that on the spatial domain, while it outperforms the benchmark of the fitted trend. The two least squares estimation on the frequency and spatial domains shares almost equivalent results, although the one on the spatial domain is slightly better. It may come from the approximation to evaluate the out-of-sample forecats in (7) that gives advantage to the estimation on the spatial domain in (6) over the one on the frequency domain in (5).

5.2 COVID-19 outbreak in Japan

This section applies the convolutional regression in (1) with CAR(1) kernels to analysis of COVID-19 outbreak in Japan in relation with human mobilities. We aim to identify the impact of human mobility to outbreak of COVID-19.

We collected the weekly new cases of COVID-19 infections at all of 1,896 cities in Japan as dependent variables, while we evaluated weekly human mobility based on dataset provided by NTT DOCOMO, Inc., one of the largest Japanese mobile phone companies as an independent variable. More precisely, the dependent variable is the covid19 infection rate, which is, for n_{ti} and N_{ti}, weekly new cases and population at a city i and a week t, defined by,

$$covid19_t(s_i) = \frac{n_{ti}}{N_{ti}} \times 10,000.$$

NTT Docomo, inc. provides spatio-temporal dataset that counts numbers of people in 500 meter meshes at every one hour period all over Japan. We aggregate it into city level that counts the weekly number of people who enter a city from outside the city. For m_{ti}, weekly number of people who enter a city i at s_i from outside at a week t, and define the independent variable by

$$hmobil_t(s_i) = \frac{m_{ti}}{N_{ti}}.$$

In addition, we collected weekly total precipitation at 936 weather stations in Japan as one more possible independent variable.

All the variables have longitudes and latitudes to indicate their locations. We transformed the coordinate of latitudes and longitudes into Cartesian coordinate based on global distance with an unit of 50 km. We took the dataset collected weekly for 86 weeks from Feb 12th, 2020 to Sep. 30th, 2021, and the last ten weeks were separated as the out-of -samples.

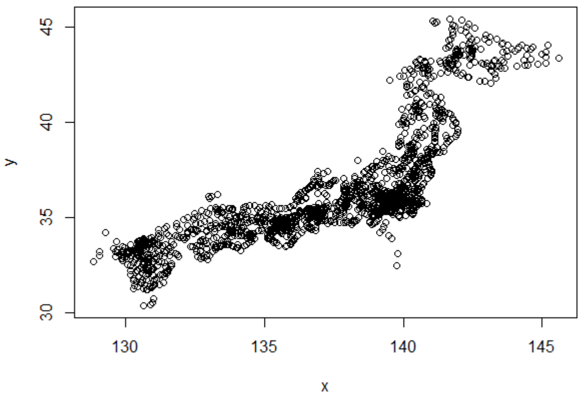

Figure 3: Locations of 1,896 Cities in Japan

We fitted the convolutional regression that includes lagged terms of dependent variables as

$$
\begin{aligned}
& covid19_t(s) \\
&= \beta_1 \int \phi_0(s-u;\delta_1) covid19_{t-1}(u)du + \beta_2 \int \phi_0(s-u;\delta_2) covid19_{t-2}(u)du \\
&+ \beta_3 \int \phi_0(s-u;\delta_3) hmobil_{t-1}(u)du + \beta_4 \int \phi_0(s-u;\delta_4) hmobil_{t-2}(u)du \\
&+ \beta_5 \int \phi_0(s-u;\delta_5) ppt_{t-2}(u)du + \varepsilon_t(s), \qquad (15)
\end{aligned}
$$

for the CAR(1) kernel $\phi_0(\cdot;\delta)$ in (10).

We estimated the parameters by minimizing (5) with (9), where we took $A = [0,40] \times [0,40]$ to construct the periodogram on the mesh points $(2\pi i/A_1, 2\pi j/A_2), i, j \neq 0$ in $D = \{\omega \in \mathbb{R}^2, ||\omega|| < K\}$, where K was adjusted to make the number of the elements be 1800. The estimation results are shown in Table 4.

Then we constructed the forecasts by (15) with (7) with the estimated parameters to compare the mean squared errors with those of the benchmarks. The MSEs evaluated by, for $v_{kt}, k = 1, \ldots, n_t, t = 77, \ldots, 86$, the out-of-sample period,

$$
\frac{1}{n_{77} + \cdots + n_{86}} \sum_{t=77}^{86} \sum_{k=1}^{n_t} \Big(covid19_t(v_{kt}) - cov\hat{i}d19_t(v_{kt}) \Big)^2
$$

are shown in Table 5.

We find from Table 4 that covid-19 is a strongly autocorrelated data by the estimators for the first and second lagged autoregressive terms. Hmobil

variable	param.	LSE by (5)	se
$covid_{-1}$	β_1	0.99	(0.021)
	δ_1	0.14	(0.0078)
$covid_{-2}$	β_2	−0.39	(0.027)
	δ_2	0.45	(0.027)
$hmobil_{-1}$	β_3	−0.50	(0.062)
	δ_3	0.11	(0.052)
$hmobil_{-2}$	β_4	0.59	(0.065)
	δ_4	0.12	(0.043)
ppt_{-2}	β_5	−0.0031	(0.00092)
	δ_5	0.13	(0.19)

Table 4: The estimated parameters for (15) with their standard errors, when we took $A = [0, 40] \times [0, 40]$ to construct the periodogram on the mesh points $(2\pi i/A_1, 2\pi j/A_2), i, j \neq 0$ in $D = \{\omega \in \mathbb{R}^2, ||\omega|| < K\}$, where K was adjusted to make the number of the elements be 1800.

is negatively correlated with covid-19 at the first lag, while it is positively related at the second lag, when we control the first and second lagged covid-19. The negative correlation with the first lag means that human mobility is strictly controlled by local or central government. Namely the government predicts the tendency of increasing or decreasing new cases and fixes the policies to control the mobilities based on the prediction, which can produce the negative correlation at the first lag. On the other hand, the second lagged human mobility positively related with covid-19, which demonstrates the well known result that it takes around two weeks for a man to be infected with the virus. Finally we find that the precipitation at the second lag correlates covid-19 negatively. From the fitted models that are not shown here, it is seen that the precipitation at any other lags is not significant. Although it is not clear which causes the negative correlation in two weeks, restricted human mobility by rain or effects of humidity, it is interesting to detect the negative relationship by the convolutional regression.

Table 5 indicates that the least square estimation on frequency domain has larger MSEs than that on the spatial domain in terms of both in-sample and out-of-sample comparisons, although they should be theoretically equivalent. The difference here is even greater than that in the US weather data case in Table mse. The greater difference of MSE as well as the estimation between the two estimation may come from the difference of the sample sizes. There are around 6,000 weather stations in the first example, while there are around 1,800 cities recorded for weekly new cases of infections. Frequency domain estimation requires more sample sizes than that on the spatial domain to attain the equivalent approximation accuracy.

model	forecasting MSE in-sample	out-of-sample
CAR(1) by (5)	1.82	12.7
CAR(1) by (6)	1.71	11.8
trend fit by (13)	2.45	31.3

Table 5: Forecasting MSE for in-samples (77 weeks in Feb. 2020-Jul. 2021) and out-of-samples (8 weeks in Aug. -Sep. 2021) in Covid-19 new cases over 1,760 cities in Japan.

6 Discussion

This paper tries a parametric convolutional regression for big spatial data on huge scales at discrete time points. Applying Fourier transform to the regression, we propose the least squares estimation on the frequency domain. We have found several features of the convolutional regression for big spatial data.

First, our approach can handle a regression among big spatial data on national or continental scale. Second, our approach does not require stationarity across time but do across space, while existing approaches on functional PCA need not stationarity across space but do across time. Non-stationary behaviours including unit root across time or even one time point observation suffices to estimate parametric convolution kernels consistently. Stationarity across space necessary for our approach to work is the price for no restrictions across time. As a result, our regression is free from spurious regression problem indicated by Granger and Newbold (1974), which is well demonstrated in the simulation studies. Thirdly, our approach allows irregularly spaced data over space and irregularity can change with time varying sample sizes across time, which is common in collecting big spatial data. The discrete Fourier transform works critically to make it possible to estimate properly under irregular sampling across space. Finally, the leat squares estimation on the frequency domain is validated asymptotically under the so called mixed asymptotics. Since the consistent asymptotic variance matrix estimator is easily constructed, statistical inference for convolutional regression becomes possible. The t test in the empirical study of Covid-19 analysis in relation with human mobility data collected on national scale in Japan detects the significant relation statistically that two week lagged human mobility associates positively current week Covid-19 new cases.

7 Proofs

We define $L(\theta)$ by normalizing $Q(\theta)$ in (5) as

$$
\begin{aligned}
L(\theta) &= (2\pi)^2 Q(\theta) \\
&= \frac{(2\pi)^2}{|A|} \sum_{\omega_f \in D} \sum_{t=1}^{T} \left\{ I_{t,yy}(\omega_f) - 2\tilde{\phi}(\omega_f;\theta) I_{t,xy}(\omega_f) + \tilde{\phi}(\omega_f;\theta) I_{t,xx}(\omega_f) \tilde{\phi}'(\omega_f;\theta) \right\},
\end{aligned}
$$

where $\tilde{\phi}(\omega;\theta) = (\tilde{\phi}_1(\omega;\theta), \ldots, \tilde{\phi}_p(\omega,\theta))$ and

$$
\begin{aligned}
I_{t,xx}(\omega) &= |A| \hat{x}_t(\omega) \overline{\hat{x}_t(\omega)}', \\
I_{t,xy}(\omega) &= |A| \hat{x}_t(\omega) \overline{\hat{y}_t(\omega)}, \\
I_{t,yy}(\omega) &= |A| \hat{y}_t(\omega) \overline{\hat{y}_t(\omega)},
\end{aligned}
$$

the periodograms.

7.1 Proof of Theorem 1

Let $\theta_1 \in \Theta$ be a parameter that is not equal to θ_0. By applying Lemmas 1 and 2 to the periodograms with the relations,

$$
\begin{aligned}
& f_{t,yy}(\omega) = \tilde{\phi}(\omega;\theta_0) f_{t,xx}(\omega) \tilde{\phi}'(\omega;\theta_0) + f_{t,\varepsilon\varepsilon}(\omega), t = 1, \ldots, T, \\
& f_{t,xy}(\omega) = f_{t,xx}(\omega) \tilde{\phi}'(\omega;\theta_0), \\
& f_{t,x\varepsilon}(\omega) = 0,
\end{aligned}
$$

we have

$$
\begin{aligned}
L(\theta_1) \to & \tau_g \sum_{t=1}^{T} \int_D \left\{ \left(\tilde{\phi}(\omega;\theta_1) - \tilde{\phi}(\omega;\theta_0) \right) f_{t,xx}(\omega) \left(\tilde{\phi}(\omega;\theta_1) - \tilde{\phi}(\omega;\theta_0) \right)' + f_{t,\varepsilon\varepsilon}(\omega) \right\} d\omega \\
& := L_\infty(\theta_1),
\end{aligned}
$$

say, in probability as k tends to be infinity in the asymptotic regime of C1. Since $f_{t,xx}(\omega)$ is positive definite,

$$
\begin{aligned}
& L_\infty(\theta_1) - L_\infty(\theta_0) = \\
& \qquad \tau_g \sum_{t=1}^{T} \int_D \left(\tilde{\phi}(\omega;\theta_1) - \tilde{\phi}(\omega;\theta_0) \right) f_{t,xx}(\omega) \left(\tilde{\phi}(\omega;\theta_1) - \tilde{\phi}(\omega;\theta_0) \right)' d\omega \\
& \qquad > 0.
\end{aligned}
$$

It follows that, for any positive constant $K(\theta_0, \theta_1)$ that is smaller than $L_\infty(\theta_1) - L_\infty(\theta_0)$,

$$
\lim_{k \to \infty} P\left(L(\theta_0) - L(\theta_1) < -K(\theta_0, \theta_1) \right) = 1.
$$

Let $M_t(\omega)$ be the maximum eigenvalue of the periodogram matrix

$$I_t(\omega) = \begin{pmatrix} I_{t,yy}(\omega) & \overline{I_{t,xy}(\omega)}' \\ I_{t,xy}(\omega) & I_{t,xx}(\omega) \end{pmatrix}, \omega \in D, t = 1, \ldots, T.$$

Then for any θ_1 and θ_2 that satisfy $|\theta_2 - \theta_1| < \delta$,

$$|L(\theta_2) - L(\theta_1)| < C\delta \sup_{\omega \in D} \max_t M_t(\omega) = H_\delta,$$

say. It is seen that there exists a $\delta > 0$ such that

$$\lim_{k \to \infty} P(H_\delta < K(\theta_0, \theta_1)) = 1.$$

Applying Lemma 2 in Walker (1964), we have the consistency.

7.2 Proof of Theorem 2

Applying Taylor series expansion to $\frac{\partial L(\hat{\theta})}{\partial \theta}$ at θ_0, we have

$$0 = \frac{\partial L(\hat{\theta})}{\partial \theta} = \frac{\partial L(\theta_0)}{\partial \theta} + \frac{\partial^2 L(\theta^*)}{\partial \theta \partial \theta'}(\hat{\theta} - \theta_0),$$

where θ^* is the mean value between θ_0 and $\hat{\theta}$. Hence

$$\sqrt{|A|}\left(\hat{\theta} - \theta_0\right) = \left\{\frac{\partial^2 L(\theta^*)}{\partial \theta \partial \theta'}\right\}^{-1} \sqrt{|A|}\left\{-\frac{\partial L(\theta_0)}{\partial \theta}\right\}.$$

The first factor, which is the Hessian matrix, is evaluated as,

$$2\frac{(2\pi)^2}{|A|} \sum_{\omega_k \in D} \frac{\partial \tilde{\phi}(\omega_k; \theta^*)}{\partial \theta'} \left\{\sum_t I_{t,xx}(\omega_k)\right\} \frac{\partial \tilde{\phi}'(\omega_k; \theta^*)}{\partial \theta}$$
$$-2\frac{(2\pi)^2}{|A|} \sum_t \sum_{\omega_f \in D} \sum_{a=1}^{p} \left\{I_{t,yx_a}(\omega_f) - I_{t,x_a x}(\omega_f)\tilde{\phi}(\omega_f; \theta^*)\right\} \frac{\partial^2 \tilde{\phi}_a(\omega_f; \theta^*)}{\partial \theta \partial \theta'},$$

which converges in probability by Lemmas 1 and 2 to

$$2\tau_g (2\pi)^2 \Omega, \tag{16}$$

since θ^* converges to θ_0 under the consistency, and $f_{t,yx}(\omega) = f_{t,xx}(\omega)\tilde{\phi}'(\omega; \theta_0)$.

The second factor, which is the score vector, is evaluated as

$$-2\frac{\sqrt{|A|}(2\pi)^2}{|A|} \sum_{\omega_f \in D} \sum_{t=1}^{T} \frac{\partial \tilde{\phi}(\omega_f; \theta_0)}{\partial \theta} |A| \hat{x}_t(\omega_f) \left\{\overline{\hat{y}_t(\omega_f) - \tilde{\phi}(\omega_f; \theta_0)\hat{x}_t(\omega_f)}\right\}$$

which is, by Lemma 2, equal to

$$-2\frac{\sqrt{|A|}(2\pi)^2}{|A|}\sum_{\omega_f\in D}\frac{\partial\tilde{\phi}(\omega_f;\theta_0)}{\partial\theta}\left\{\sum_{t=1}^{T}I_{t,x\varepsilon}(\omega_f)\right\}+J, \tag{17}$$

where

$$I_{t,x\varepsilon}(\omega)=|A|\hat{x}_t(\omega)\overline{\hat{\varepsilon}_t(\omega)},$$
$$\hat{\varepsilon}_t(\omega)=\frac{1}{n_{0t}}\sum_{j=1}^{n_{0t}}\varepsilon_t(s_{0tj})\exp(-i\omega' s_{0tj}),$$

and $E|J|$ is bounded by

$$C\max_{\omega\in D}\max_{1\le t\le T}\max_{1\le a\le p}$$
$$\times\sqrt{|A|^{1/2}E\left|\frac{|A|^{1/2}}{n_{0t}}\sum_{j=1}^{n_{0t}}x_{ta}(s_{0tj}-u)e^{-i(s_{0tj}-u)'\omega}-\frac{|A|^{1/2}}{n_{at}}\sum_{j=1}^{n_{at}}x_{ta}(s_{atj})e^{-is'_{atj}\omega}\right|^2}$$
$$=o(1),$$

by Lemma 1. Hence it is equivalent to show the asymptotic distribution of

$$K=2\sqrt{|A|}\int_D\frac{\partial\tilde{\phi}(\omega;\theta_0)}{\partial\theta}\left\{\sum_{t=1}^{T}I_{t,x\varepsilon}(\omega)\right\}d\omega,$$

by applying Lemma 3 to the first term.

We shall show in Lemmas 4, 5 and 6 that

$$E(K_a)\to 0, a=1,\dots,q,$$
$$E(K_a\overline{K}_b)\to 4(2\pi)^6\tau_{g^2}\Sigma_{ab}, a,b=1,\dots,q,$$
$$cum\,(K_{a_1},\dots,K_{a_r}))\to 0,\ \text{for } 1\le a_1,\dots,a_r\le q, r\ge 3,$$

respectively, as k tends to ∞, which proves the asymptotic normality in Theorem 2 in combinations with (16).

8 Lemmas

For the unified treatment of the periodograms, we define, for $t=1,\dots,T$,

$$z_{t0}(s)=y_t(s),$$
$$z_{ta}(s)=x_{ta}(s), a=1,\dots,p,$$

and, re-define the periodogram and spectral density matrix of z by, for $a,b=0,1,\dots,p$,

$$I_{t,ab}(\omega)=|A|\hat{z}_{ta}(\omega)\overline{\hat{z}_{tb}(\omega)},$$
$$f_{t,ab}(\omega)=f_{t,z,ab}(\omega), a,b=0,1,\dots,p,$$

Lemma 1.

$$EI_{t,ab}(\omega) = (2\pi)^2 \tau_g f_{t,ab}(\omega) + O\left(\frac{n_{tab}|A|}{n_{ta}n_{tb}} + A_1^{-2} + A_2^{-2}\right), a, b = 0, 1, \ldots, p,$$

where n_{ta}, n_{tb} and n_{tab} are the number of elements in S_{ta}, S_{tb} and $S_{ta} \cap S_{tb}$, respectively, and

$$\tau_g = \int_{[0,1]^2} |g(x)|^2 dx.$$

Proof. This is an immediate consequence of Lemma 3 in Matsuda and Yajima (2009). □

Lemma 2. *For a square integrable function $\psi(\omega), \omega \in \mathbb{R}^2$,*

$$var\left\{\frac{(2\pi)^2}{|A|} \sum_{\omega_j \in D} I_{t,ab}(\omega_j)\psi(\omega_j)\right\} = O\left(|A|^{-1}\right), a, b = 0, 1, \ldots, p.$$

Proof. Let

$$\hat{\psi}(s) = \frac{(2\pi)^2}{|A|} \sum_{\omega_j \in D} \psi(\omega_j) e^{-i\omega_j' s}, s \in A = [0, A_1] \times [0, A_2],$$

which is extended periodically to $[-A_1, A_1] \times [-A_2, A_2]$. Then the object for the variance is evaluated as

$$\frac{|A|^{3/2}}{n_{ta}n_{tb}} \sum_{c=1}^{n_{ta}} \sum_{d=1}^{n_{tb}} z_{ta}(s_{atc}) z_b(s_{btd}) \hat{\psi}(s_{atc} - s_{btd}).$$

The variance is given by

$$\begin{aligned}
& E\frac{|A|^3}{n_{ta}^2 n_{tb}^2} \sum_{c_1}\sum_{d_1}\sum_{c_2}\sum_{d_2}\Big(cum\left(z_{ta}(s_{atc_1}), z_{tb}(s_{btd_1}), z_{ta}(s_{atc_2}), z_{tb}(s_{btd_2})\right) \\
& + \gamma_{t,aa}(s_{atc_1} - s_{atc_2})\gamma_{t,bb}(s_{btd_1} - s_{btd_2}) + \gamma_{t,ab}(s_{atc_1} - s_{btd_2})\gamma_{t,ba}(s_{btd_1} - s_{atc_2})\Big) \\
& \times \hat{\psi}(s_{atc_1} - s_{btd_1})\hat{\psi}(s_{atc_2} - s_{btd_2}) \\
= & \int_A \int_A \int_A \int_A \Big(cum(z_{ta}(u_1), z_{tb}(v_1), z_{ta}(u_2), z_{tb}(v_2)) \\
& + \gamma_{t,aa}(u_1 - u_2)\gamma_{t,bb}(v_1 - v_2) + \gamma_{t,ab}(u_1 - v_2)\gamma_{t,ba}(v_1 - u_2)\Big) \\
& \times \hat{\psi}(u_1 - v_1)\hat{\psi}(u_2 - v_2)|A|^{-1} g(u_1/A)g(v_1/A)g(u_2/A)g(v_2/A) du_1 dv_1 du_2 dv_2 + o(1).
\end{aligned}$$

The first term is, by expressing the cumulant term with the cumulant spectrum:

$$f_{t,abab}(\omega_1,\omega_2,\omega_3) = \sum_{e,f,g,h=1}^{m} \kappa_{t,efgh}\tilde{G}_{tae}(\omega_1)\tilde{G}_{tbf}(\omega_2)\tilde{G}_{tag}(\omega_3)\overline{\tilde{G}_{tbh}(\omega_1+\omega_2+\omega_3)}, \tag{18}$$

given by

$$\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} f_{t,abab}(\omega_1,\omega_2,\omega_3)e^{i\omega_1'(u_1-v_2)}e^{i\omega_2'(v_1-v_2)}e^{i\omega_3'(u_2-v_2)}d\omega_1 d\omega_2 d\omega_3$$
$$|A|^{-1}\int_A\int_A\int_A\int_A \hat{\psi}(u_1-v_1)\hat{\psi}(u_2-v_2)g(u_1/A)g(v_1/A)g(u_2)/Ag(v_2/A)du_1dv_1du_2dv_2,$$

which is, by Schwarz inequality, bounded by

$$\prod_{j=1}^{2}\sqrt{\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}|f_{t,abab}(\omega_1,\omega_2,\omega_3)|P_j d\omega_1 d\omega_2 d\omega_3}, \tag{19}$$

for

$$P_1 = |A|^{-1}\left|\int_A\int_A \hat{\psi}(u_1-v_1)g(u_1/A)g(v_1/A)e^{i\omega_1' u_1}e^{i\omega_2' v_1}du_1dv_1\right|^2,$$
$$P_2 = |A|^{-1}\left|\int_A\int_A \hat{\psi}(u_2-v_2)g(u_2/A)g(v_2/A)e^{-i(\omega_1+\omega_2)' v_2}e^{i\omega_3'(u_2-v_2)}du_2dv_2\right|^2.$$

By applying Perseval's equality to both terms, (19) is bounded by

$$C\int_{-A_1}^{A_1}\int_{-A_2}^{A_2}\left|\hat{\psi}(s)\right|^2 ds, \tag{20}$$

which is evaluated as

$$4C|A|\frac{(2\pi)^4}{|A|^2}\sum_{\omega_j\in D}|\psi(\omega_j)|^2 < C'\int_D|\psi(\omega)|^2 d\omega = O(1).$$

Also the second and third terms in the variance are bounded by a constant with the same argument, which completes the proof. □

Lemma 3. *For a square integrable function $\psi(\omega), \omega \in \mathbb{R}^2$,*

$$\frac{\sqrt{|A|}(2\pi)^2}{|A|}\sum_{j\in J_D} I_{t,x_a\varepsilon}(\omega_j)\psi(\omega_j) - \sqrt{|A|}\int_D I_{t,x_a\varepsilon}(\omega)\psi(\omega)d\omega$$
$$= o_p(1), a = 1,\dots,p, t = 1,\dots,T.$$

Proof. Let

$$\hat{\psi}(s) = \frac{(2\pi)^2}{|A|} \sum_{\omega_j \in D} \psi(\omega_j) e^{-i\omega_j' s}, s \in A = [0, A_1] \times [0, A_2],$$

$$\tilde{\psi}(s) = \int_D \psi(\omega) e^{-i\omega' s} d\omega, s \in \mathbb{R}^2,$$

the first one of which is extended periodically to $[-A_1, A_1] \times [-A_2, A_2]$. For $\delta(s) = \hat{\psi}(s) - \tilde{\psi}(s)$, the difference is evaluated as

$$L = \frac{|A|^{3/2}}{n_{ta} n_{t0}} \sum_{c=1}^{n_{ta}} \sum_{d=1}^{n_{t0}} x_a(s_{tac}) \varepsilon(s_{t0d}) \delta(s_{tac} - s_{t0d}).$$

$E|L|^2$, which is similarly evaluated till (20) in Lemma 2, is bounded by

$$C \int_{-A_1}^{A_1} \int_{-A_2}^{A_2} |\delta(s)|^2 \, ds. \tag{21}$$

Notice that $\hat{\psi}(s), \tilde{\psi}(s)$ are square integrable, since

$$\int_0^{A_1} \int_0^{A_2} |\hat{\psi}(s)|^2 ds = |A| (2\pi)^4 |A|^{-2} \sum_{\omega_j \in D} |\psi(\omega_j)|^2 < C \int_D |\psi(\omega)|^2 d\omega < \infty,$$

$$\int_{\mathbb{R}^2} |\tilde{\psi}(s)|^2 ds = (2\pi)^2 \int_D |\psi(\omega)|^2 d\omega < \infty.$$

It follows that, for any $\varepsilon > 0$, there exists a compact set $B_M = [-M_1, M_1] \times [-M_2, M_2] \subset [-A_1, A_1] \times [-A_2, A_2]$ such that

$$\int_{-A_1}^{A_1} \int_{-A_2}^{A_2} \left| \hat{\psi}(s) - \hat{\psi}(s) I_{B_M}(s) \right|^2 ds < \varepsilon,$$

$$\int_{\mathbb{R}^2} \left| \tilde{\psi}(s) - \tilde{\psi}(s) I_{B_M}(s) \right|^2 ds < \varepsilon.$$

Then (21) is bounded by

$$\begin{aligned} & C \left\{ \int_{-A_1}^{A_1} \int_{-A_2}^{A_2} \left| \hat{\psi}(s) - \hat{\psi}(s) I_{B_M}(s) \right|^2 ds + \int_{B_M} \left| \hat{\psi}(s) - \tilde{\psi}(s) \right|^2 ds \right. \\ & \quad \left. + \int_{\mathbb{R}^2} \left| \tilde{\psi}(s) I_{B_M}(s) - \tilde{\psi}(s) \right|^2 ds \right\} \\ & \quad < C' \varepsilon, \end{aligned}$$

which completes the proof. □

Lemma 4.

$$E(K_a) = O\left(A_1^{-3/2} A_2^{1/2} + A_1^{1/2} A_2^{-3/2} + \sum_{t=1}^{T} \frac{|A|^{3/2}}{n_{ta}} \right), a = 1, \ldots, q.$$

Proof. It follows immediately from Lemma 3 in Matsuda and Yajima (2009) and $f_{t,\varepsilon x}(\omega) = 0$. □

Lemma 5.

$$E(K_a \overline{K}_b) \to 4(2\pi)^6 \tau_{g^2} \Sigma_{ab}, a, b = 1, \ldots, q,$$

Proof. Re-express K_a by

$$\sum_{t=1}^{T} \sum_{c=1}^{p} \frac{|A|^{3/2}}{n_{0t} n_{ct}} \sum_{j=1}^{n_{ct}} \sum_{l=1}^{n_{0t}} X_{tc}(s_{ctj}) \varepsilon_t(s_{0tl}) \tilde{\psi}_{ac}(s_{ctj} - s_{0tl}), \tag{22}$$

where

$$\psi_{ac}(\omega) = \frac{\partial \tilde{\phi}_c(\omega; \theta_0)}{\partial \theta_a},$$
$$\tilde{\psi}_{ac}(s) = \int_D \psi_{ac}(\omega) e^{-i\omega' s} d\omega,$$

First, $\tilde{\psi}_{ac}(s)$ in K_a defined in (22) may be replaced with

$$\tilde{\psi}_{ac}^M(s) = \tilde{\psi}_{ac}(s) I_{B_M}(s),$$

for a sufficiently large compact set $B_M = [-M_1, M_1] \times [-M_2, M_2]$, since the variance of the difference between them may be made arbitrary small by following the argument till (20) in Lemma 2. Then $EK_a \bar{K}_b$ replaced with $\tilde{\psi}_{ac}^M(s)$ and $\tilde{\psi}_{bc}^M(s)$ is evaluated as

$$\sum_{t_1,t_2=1}^{T} \sum_{c_1,c_2=1}^{p} |A|^{-1} \int_A \int_A \int_A \int_A \gamma_{x,t_1 t_2, c_1 c_2}(u_1 - u_2) \gamma_{\varepsilon, t_1 t_2}(v_1 - v_2) \tilde{\psi}_{ac_1}^M(u_1 - v_1) \tilde{\psi}_{bc_2}^M(u_2 - v_2)$$
$$\times g(u_1/A) g(v_1/A) g(u_2/A) g(v2/A) du_1 dv_1 du_2 dv_2 + o(1).$$

The first term is evaluated as

$$\sum_{t_1,t_2=1}^{T} \sum_{c_1,c_2=1}^{p} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} f_{x,t_1 t_2,,c_1 c_2}(\omega_1) f_{\varepsilon, t_1 t_2}(\omega_2) |A|^{-1} P d\omega_1 d\omega_2$$

for

$$P = \int_A \int_A \int_A \int_A \tilde{\psi}_{ac1}^M(u_1 - v_1) \tilde{\psi}_{bc_2}^M(u_2 - v_2) e^{i\omega_1'(u_1 - u_2)} e^{i\omega_2'(v_1 - v_2)}$$
$$\times g(u_1/A) g(v_1/A) g(u_2/A) g(v_2/A) du_1 dv_1 du_2 dv_2.$$

By change of variables by $u_1 - v_1 = l_1, u_2 - v_2 = l_2$ and the compactness of the supports of $\tilde{\phi}^M$, $|A|^{-1} P$ is evaluated as

$$(2\pi)^4 \psi_{ac_1}^M(\omega_1) \psi_{bc_2}^M(\omega_1) F_{g^2}(\omega_1 + \omega_2) + o(1),$$

where

$$F_{g^2}(\omega) = |A|^{-1} \left| \int_A g^2(u/A) e^{i\omega' u} du \right|^2,$$
$$\psi_{ac_1}^M(\omega) = (2\pi)^{-2} \int_{B_M} \tilde{\psi}_{ac_1}^M(s) e^{i\omega' s} ds.$$

It follows by Lemma 1(c) in Matsuda and Yajima (2009) that the first term converges to

$$\sum_{t_1,t_2=1}^{T} \sum_{c_1,c_2=1}^{p} (2\pi)^6 \tau_{g^2} \int_{\mathbb{R}^2} f_{x.t_1t_2,c_1c_2}(\omega) f_{\varepsilon,t_1t_2}(\omega) \psi_{ac_1}^M(\omega) \psi_{bc_2}^M(\omega) d\omega.$$

We have the result arbitrary close to $(2\pi)^6 \tau_{g^2} \Sigma_{ab}$ by taking B_M large. □

Lemma 6. *For* $r \geq 3$,

$$cum(K_{a_1}, \ldots, K_{a_r}) = O\left(|A|^{-r/2+1}\right).$$

Proof. First, $\tilde{\psi}_{ac}(s)$ in K_a defined in (22) may be replaced with

$$\tilde{\psi}_{ac}^M(s) = \tilde{\psi}_{ac}(s) I_{B_M}(s), a = 1, \ldots, q,$$

for a sufficiently large compact set $B_M = [-M_1, M_1] \times [-M_2, M_2]$, since the variance of the difference between them may be made arbitrary small by following the argument till (20) in Lemma 2.

The cumulant for the one replaced with $\tilde{\psi}_{ac}^M(s)$ is evaluated as

$$cum(K_{a_1}, \ldots, K_{a_r}) = \sum_{t_1,\ldots,t_r=1}^{T} \sum_{c_1,d_1=1}^{q} \cdots \sum_{c_r,d_r=1}^{q} |A|^{-r/2}$$
$$\times \int_A \cdots \int_A cum(x_{t_1,c_1}(u_1)\varepsilon_{t_1}(v_1), \ldots, x_{t_r,c_r}(u_r)\varepsilon_{t_r}(v_r))$$
$$\times \prod_{j=1}^{r} \tilde{\psi}_{a_jc_j}^M(u_j - v_j) g(u_j/A) g(v_j/A) du_j dv_j + o(1),$$

in which the cumulant is evaluated as

$$cum(x_{t_1,c_1}(u_1), \ldots, x_{t_r,c_r}(u_r)) \times cum(\varepsilon_{t_1}(v_1), \ldots, \varepsilon_{t_r}(v_r))$$
$$= \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} f_{x,t_1\ldots t_r,c_1\cdots c_r}(\omega_1, \ldots, \omega_{r-1}) \prod_{j=1}^{r-1} e^{i\omega_j'(u_j-u_r)} d\omega_1 \cdots d\omega_{r-1}$$
$$\times \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} f_{\varepsilon,t_1\ldots t_r}(\lambda_1, \ldots, \lambda_{r-1}) \prod_{j=1}^{r-1} e^{i\omega_j'(v_j-v_r)} d\lambda_1 \cdots d\lambda_{r-1}.$$

By replacing the cumulant term with the spectrum expression, the summand of the corresponding term is given by

$$
\begin{aligned}
&|A|^{-r/2}\int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} f_{x,t_1\ldots t_r,c_1\cdots c_r}(\omega_1,\ldots,\omega_{r-1})d\omega_1\cdots\omega_{r-1}\\
&\times\int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} f_{\varepsilon,t_1\ldots t_r}(\lambda_1,\ldots,\lambda_{r-1})d\lambda_1\cdots\lambda_{r-1}\\
&\times\prod_{j=1}^{r-1}\int_A\int_A \tilde{\psi}^M_{a_jc_j}(u_j-v_j)e^{i\omega_j' u_j}e^{i\lambda_j' v_j}g(u_j/A)g(v_j/A)du_jdv_j\\
&\times\int_A\int_A \tilde{\psi}^M_{a_rc_r}(u_r-v_r)e^{-i(\omega_1+\cdots+\omega_{r-1})' u_r}e^{-i(\lambda_1+\cdots+\lambda_{r-1})' v_r}g(u_r/A)g(v_r/A)du_rdv_r\\
=&C|A|^{-r/2}\int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} f_{x,t_1\ldots t_r,c_1\cdots c_r}(\omega_1,\ldots,\omega_{r-1})d\omega_1\cdots\omega_{r-1}\\
&\times\int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} f_{\varepsilon,t_1\ldots t_r}(\lambda_1,\ldots,\lambda_{r-1})d\lambda_1\cdots\lambda_{r-1}\times\prod_{j=1}^{r-1}D_2(\omega_j+\lambda_j)\psi^M_{a_jc_j}(\omega_j)\\
&\times D_2(-\omega_1-\lambda_1-\cdots-\omega_{r-1}-\lambda_{r-1})\psi^M_{a_jc_j}(-\omega_1-\cdots-\omega_{r-1})+o(1),
\end{aligned}
$$

where

$$D_2(\omega)=\int_A g^2(u/A)e^{-\omega' u}du.$$

We find it to be $O(|A|^{-r/2+1})$ by Lemma 2 in Matsuda and Yajima (2009). □

References

[1] Benatia, D., Carrasco, M. and Florens, J. (2017) Functional linear regression with functional response, *Journal of Econometrics*, 201, 269-291.

[2] Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer-Verlag, New York.

[3] Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*, 2nd edn. Springer-Verlag.

[4] Brockwell, P. J. and Matsuda, Y. (2017). Continuous auto-regressive moving average random fields on $\mathbb{R}^n$. *Journal of the Royal Statistical Society: Series B*, 79(3), 833-857.

[5] Crambes, C. and Mas, A. (2013). Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli* 19, 2627–2651.

[6] Ferraty, F., Vieu, P. (2006). Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations. Springer-Verlag, London.

[7] Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2, 111-120.

[8] Horvath, L., Kokoszka, P. (2012). Inference for Functional Data with Applications. Springer, New-York.

[9] Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models *Annals of Statistics*, .25, 1, 105-37.

[10] Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis.* CRC Press, Boca Raton.

[11] Liu, X., Xiao, H. and Chen, R. (2016). Convolutional autoregressive models for functional time series. *Journal of Econometrics*, 194, 263-282.

[12] Matsuda, Y. and Yajima, Y. (2009). Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *Journal of the Royal Statistical Society: Series B*, 71(1), 191-217.

[13] Matsuda, Y. and Yajima, Y. (2018). Locally stationary spatio-temporal processes. *Jpn. J. Stat. Data Sci.*, 1, 41-57.

[14] Ramsay, J.O., Silverman, B.W.(2005). Functional Data Analysis, second ed. Springer, New York.

[15] Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23, 1630-1661.

[16] Stein, M. L. (1999). *Interpolation of Spatial Data.* Springer, New York.

[17] Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of Brownian Motion, *Physical Review*, **36**, 823-841.

[18] Walker, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *Journal of the Australian Mathematical Society*, 4, 363-384.

[19] Yao, F., Müller, H.G., Wang, J.L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* 33 (6), 2873–2903.