CARF Working Paper

CARF-F-517

Backward Smoothing for Noisy Non-stationary Time Series

Seisho Sato
Graduate School of Economics, University of Tokyo

Naoto Kunitomo
Gendai-Finance-Center, Tokyo Keizai University

July 8, 2021

CARF is presently supported by Nomura Holdings, Inc., Sumitomo Mitsui Banking Corporation, The Dai-ichi Life Insurance Company, Limited, The Norinchukin Bank, MUFG Bank, Ltd. and Ernst & Young ShinNihon LLC. This financial support enables us to issue CARF Working Papers.

Backward Smoothing for Noisy Non-stationary Time Series *

Seisho Sato †
and
Naoto Kunitomo ‡

July 8, 2021

Abstract

In this study, we investigate a new smoothing approach to estimate the hidden states of random variables and to handle multiple noisy non-stationary time series data. Kunitomo and Sato (2021) have developed a new method to solve the smoothing problem of hidden random variables, and the resulting separating information maximum likelihood (SIML) method enables the handling of multivariate non-stationary time series. We continue to investigate the filtering problem. In particular, we propose *the backward SIML smoothing method* and *the multi-step smoothing method* to address the initial value issue. The resulting filtering methods can be interpreted in the time and frequency domains.

Key Words

Non-stationary multivariate economic time series, Errors-in-variables models, SIML-backward-smoothing, Band smoothing, Multi-step smoothing, Initial value problem.

*2021-7-8. This is a revised version of the paper (SDS-15, MIMS=RSP Statistics and Data Science Series, Meiji University). The research has been supported by JSPS-Grant JP17H02513 and CARF.

†Graduate School of Economics, University of Tokyo, Bunkyo-ku,Hongo7-3-1, Tokyo 113-0033, seisho@e.u-tokyo.ac.jp

‡Gendai-Finance-Center, Tokyo Keizai University, Kokubunji-shi, Minami-cho 1-7-34, 185-8502, Tokyo, JAPAN. naoto.kunitomo@gmail.com

1. Introduction

We investigate a new smoothing method to estimate the hidden states of random variables and to handle multiple noisy non-stationary time series data. The main motivation for the study has been to handle small sample non-stationary economic time series. Kunitomo, Sato, and Kurisu (2018) have developed the separating information maximum likelihood (SIML) method for financial high-frequency time series. Subsequently, Kunitomo and Sato (2021) utilized their results to solve the smoothing or filtering problem of hidden random variables, which enables a new estimation method to handle macro-economic time series. In this study, we continue to investigate the smoothing or filtering problem; in particular, we develop the *backward* SIML smoothing method and *multi-step* SIML smoothing method to address the initial value issue in the procedure. Based on our analysis, the SIML forward and backward smoothing methods can be interpreted both in the time and frequency domains. Although some econometrician may not distinguish smoothing from filtering and the latter terminology has been sometimes used, we use *smoothing* mainly rather than *filtering* in this study.

A large body of published research exists on the use of time series analysis for macro-economic time series, which have non-stationary trend, cycle, seasonal. and measurement errors. For statistical filtering and smoothing methods, Kitagawa (2010) discussed the standard statistical methods already known, including the Kalman filtering and particle-filtering methods. Harvey and Trimbur (2008) have investigated the relation between Kalman filtering and Hodorick-Prescott filter in econometrics, for example. Although many studies have examined statistical filtering theories, we must exercise caution in analyzing non-stationary multivariate time series. Existing methods often depend on underlying distributions such as the Gaussian distributions for Kalman filtering. As some procedure essentially depends on the dimension of state variables, some difficulty may exist in extending the existing methods to high-dimensional cases, despite that the dimension is about 10. Meanwhile, we expect that our method is simple and has some advantage when handling small sample economic times series with non-stationarity and seasonality with many variables. It is because our method does not depend on specific distributions as well as the dimension of the underlying random variables. Refer to Kunitomo, Awaya, and Kurisu (2020) for a comparison of small sample properties of the maximum likelihood and SIML estimation methods for the non-stationary errors-in-variables model, and Nishimura, Sato, and Takahashi (2019) for an application of financial data smoothing. In particular, the most important feature of the present procedure is that it may be applicable to small sample time series data.

In Kunitomo and Sato (2021), an implicit assumption exists, that is, we can handle the initial value problem in smoothing or filtering. However, in non-stationary time series, the initial value of state estimate may play a crucial role in the resulting estimates of unobservable state vectors; thus we need to investigate this problem in a systematic manner. The present study mainly aims to resolve this problem in a general manner. In this study we shall demonstrate that backward smoothing and iterative smoothing procedures can be developed, and we obtain their convergence. However, some related issues arise, and we develop the multi-step and band-smoothing procedures, which are new in this field. It seems that they are related to the general problem in the analysis of non-stationary time series data.

The reminder of this paper is organized as follows. In Section 2, we explain the non-stationary errors-in-variables model and the SIML method. In Section 3, we develop the SIML-filtering methods, including the forward, backward, and multi-step smoothing procedures. Furthermore, we present a theoretical result of convergence of the smoothing or filtering method for the initial value problem and then discuss the evaluation criteria. In Section 4, we discuss generalizations of the non-stationary errors-in-variables model and the mathematical interpretation of our procedure. In Section 5, we provide some numerical examples and in Section 6, conclusions are drawn. Some details of mathematical derivations and figures are given in the Appendix.

2. Non-stationary Errors-in-variables models

Let y_{ji} be the $i-$th observation of the $j-$th time series at i for $i = 1, \cdots, n; j = 1, \cdots, p$. We set $\mathbf{y}_i = (y_{1i}, \cdots, y_{pi})^{'}$ as a $p \times 1$ vector and $\mathbf{Y}_n = (\mathbf{y}_i^{'})\ (= (y_{ij}))$ as an $n \times p$ matrix of observations, and we denote $\mathbf{y}_0$ as the initial $p \times 1$ vector, which is assumed to be observable. Furthermore, we attempt to estimate the underlying non-stationary trends when the nonstationary state vector $\mathbf{x}_i\ (= (x_{ji}))\ (i = 0, 1, \cdots, n)$, and the vector of noise component $\mathbf{v}_i^{'} = (v_{1i}, \cdots, v_{pi})$ are mutually independent. Then, we use the non-stationary errors-in-variables representation

$$\mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i \quad (i = 0, 1, \cdots, n), \tag{2.1}$$

where $\mathbf{x}_i\ (i = 0, 1, \cdots, n)$ is a sequence of the non-stationary I(1) process, which satisfies

$$\Delta \mathbf{x}_i = (1 - \mathcal{L})\mathbf{x}_i = \mathbf{v}_i^{(x)} \quad (i = 1, \cdots, n), \tag{2.2}$$

($\mathbf{x}_0$ as the initial vector), and $\mathbf{v}_i^{(x)}$ is a sequence of the i.i.d. random vectors with $\mathbf{E}(\mathbf{v}_i^{(x)}) = \mathbf{0}$ and $\mathbf{E}(\mathbf{v}_i^{(x)}\mathbf{v}_i^{(x)'}) = \mathbf{\Sigma}_x$. The random vector $\mathbf{v}_i\ (i = 0, 1, \cdots, n)$ is a sequence of i.i.d. random variables with $\mathbf{E}(\mathbf{v}_i) = \mathbf{0}$ and $\mathbf{E}(\mathbf{v}_i\mathbf{v}_i^{'}) = \mathbf{\Sigma}_v$.

We consider a situation wherein each pair of vectors $\Delta\mathbf{x}_i$ and $\mathbf{v}_i$ are independently, identically, and normally distributed (i.i.d.) as $N_p(\mathbf{0}, \mathbf{\Sigma}_x)$ and $N_p(\mathbf{0}, \mathbf{\Sigma}_v)$, respectively, and we have the observations of an $n \times p$ matrix $\mathbf{Y}_n = (\mathbf{y}_i^{'})$. Given the initial condition $\mathbf{y}_0$, the $np \times 1$ random vector $(\mathbf{y}_1^{'}, \cdots, \mathbf{y}_n^{'})^{'}$ follows

$$\mathrm{vec}(\mathbf{Y}_n) \sim N_{n\times p}\left(\mathbf{1}_n \cdot \mathbf{y}_0^{'}, \mathbf{I}_n \otimes \mathbf{\Sigma}_v + \mathbf{C}_n\mathbf{C}_n^{'} \otimes \mathbf{\Sigma}_x\right) , \tag{2.3}$$

where $\mathbf{1}_n^{'} = (1, \cdots, 1)$ and

$$\mathbf{C}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 0 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}_{n\times n} . \tag{2.4}$$

We use the $\mathbf{K}_n-$transformation that is from $\mathbf{Y}_n$ to $\mathbf{Z}_n$ $(= (\mathbf{z}_k^{'}))$ by

$$\mathbf{Z}_n = \mathbf{K}_n\left(\mathbf{Y}_n - \bar{\mathbf{Y}}_0\right) , \mathbf{K}_n = \mathbf{P}_n\mathbf{C}_n^{-1} , \tag{2.5}$$

where

$$\mathbf{C}_n^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}_{n\times n} , \tag{2.6}$$

and the $(k, j)-$th element of $\mathbf{P}_n = (p_{kj}^{(n)})$ is defined by

$$p_{kj}^{(n)} = \sqrt{\frac{2}{n+\frac{1}{2}}}\cos\left[\frac{2\pi}{2n+1}(k-\frac{1}{2})(j-\frac{1}{2})\right] . \tag{2.7}$$

By using the spectral decomposition, $\mathbf{C}_n^{-1}\mathbf{C}_n^{'-1} = \mathbf{P}_n\mathbf{D}_n\mathbf{P}_n$, and $\mathbf{D}_n$ is a diagonal matrix with the k-th element $d_k = 2[1-\cos(\pi(\frac{2k-1}{2n+1}))]$ $(k = 1, \cdots, n)$, and we write

$$a_{kn}^{*} \;(= d_k) \; = 4\sin^2\left[\frac{\pi}{2}\left(\frac{2k-1}{2n+1}\right)\right] \;(k = 1, \cdots, n) . \tag{2.8}$$

By taking a positive integer m_n, the SIML estimator of $\hat{\mathbf{\Sigma}}_x$ can be defined by

$$\mathbf{G}_m = \hat{\mathbf{\Sigma}}_{x,SIML} = \frac{1}{m_n}\sum_{k=1}^{m_n}\mathbf{z}_k\mathbf{z}_k^{'} . \tag{2.9}$$

Given the initial condition $\mathbf{y}_0$, the log-likelihood function, except some constants when the underlying distributions are Gaussian, can be written as

$$l_n(\boldsymbol{\theta}) = \sum_{k=1}^{n}\log|a_{kn}^{*}\mathbf{\Sigma}_v + \mathbf{\Sigma}_x|^{-1/2} - \frac{1}{2}\sum_{k=1}^{n}\mathbf{z}_k^{'}[a_{kn}^{*}\mathbf{\Sigma}_v + \mathbf{\Sigma}_x]^{-1}\mathbf{z}_k \tag{2.10}$$

and

$$(-2)l_n(\boldsymbol{\theta}) = \sum_{k=1}^{n}\log|a_{kn}^{*}\mathbf{\Sigma}_v + \mathbf{\Sigma}_x| + \sum_{k=1}^{n}\mathbf{z}_k^{'}[a_{kn}^{*}\mathbf{\Sigma}_v + \mathbf{\Sigma}_x]^{-1}\mathbf{z}_k , \tag{2.11}$$

where $\boldsymbol{\theta}$ is a vector of parameters.

The model of (2.1) and (2.2) can be generalized to the cases when we have cycle and seasonal components, and when $\mathbf{v}_t$ and $\Delta\mathbf{x}_t$ are auto-correlated (refer to Kunitomo and Sato (2021a, b)). However, in this study, we first focus on the smoothing or filtering procedure of the simple non-stationary multiple time series. We discuss briefly several extensions in Section 4.

3. SIML Smoothing and Backward Smoothing

3.1 Forward SIML Smoothing

Kunitomo and Sato (2021) investigated the general filtering procedure based on $\mathbf{K}_n$−transformation. Because the elements of the resulting $n \times p$ random matrix $\mathbf{Z}_n$ by this transformation take real values in the frequency domain, their roles are easy to understand. As $\mathbf{P}_n$ is a type of real-valued discrete Fourier transformation, vectors $\mathbf{z}_k$ $(k = 1, \cdots, n)$ in $\mathbf{Z}_n$ are asymptotically uncorrelated. We consider the partial inversion of the transformed orthogonal processes. Let an $n \times p$ matrix be

$$\hat{\mathbf{X}}_n(\mathrm{Q}) = \mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n\mathbf{P}_n\mathbf{C}_n^{-1}(\mathbf{Y}_n - \bar{\mathbf{Y}}_0) \tag{3.1}$$

and

$$\mathbf{Z}_n = \mathbf{P}_n\mathbf{C}_n^{-1}(\mathbf{Y}_n - \bar{\mathbf{Y}}_0)\ , \mathbf{Y}_n = \bar{\mathbf{Y}}_0 + \mathbf{X}_n^{(0)} + \mathbf{V}_n\ , \tag{3.2}$$

where $\mathbf{X}_n^{(0)} = (\mathbf{x}_i^{(0)'})$ and $\mathbf{V}_n = (\mathbf{v}_i^{'})$ are $n \times p$ matrices, and $\mathbf{x}_i^{(0)} = \mathbf{x}_i - \mathbf{x}_0$ $(i = 1, \cdots, n)$. We set the initial vector as $\mathbf{y}_0 = \mathbf{x}_0$.
The stochastic process $\mathbf{Z}_n$ is the orthogonal decomposition of the original time series $\mathbf{Y}_n$ in the frequency domain, and $\mathbf{Q}_n$ is an $n \times n$ filtering matrix. Because $\mathbf{Y}_n$ consists of non-stationary time series, we need a special form of transformation $\mathbf{K}_n$ in (2.5). We offer explicit form for the trend smoothing (or filtering) procedure. Let an $m \times n$ choice matrix be $\mathbf{J}_m = (\mathbf{I}_m, \mathbf{O})$, and let an $n \times p$ matrix be

$$\hat{\mathbf{X}}_n(m) = \mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n^{(m)}\mathbf{P}_n\mathbf{C}_n^{-1}(\mathbf{Y}_n - \bar{\mathbf{Y}}_0)\ , \tag{3.3}$$

and we denote an $n \times n$ matrix $\mathbf{Q}_n^{(m)} = \mathbf{J}_m^{'}\mathbf{J}_m$.
We construct an estimator of the $n \times p$ hidden state matrix $\mathbf{X}_n$ only in the lower-frequency parts by using the inverse transformation of $\mathbf{Z}_n$ and by deleting the estimated noise parts (refer to Nishimura, Sato and Takahashi (2019)). We denote the hidden trend state as

$$\mathbf{X}_n(m) = \mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n^{(m)}\mathbf{P}_n\mathbf{C}_n^{-1}\mathbf{X}_n^{(0)}\ . \tag{3.4}$$

This quantity is different from $\mathbf{X}_n$ because $\mathbf{x}_i$ $(i = 1, \cdots, n)$ in (3.1) and (3.2) contains not only the trend component of $\mathbf{y}_i$ $(i = 1, \cdots, n)$ but also the noise component in the frequency domain, which is different from the measurement noise component $\mathbf{v}_i$ $(i = 1, \cdots, n)$ of (3.1) and (3.3). We attempt to estimate the trend component of $\mathbf{x}_i$ by using (3.3) and recover the trend component of $\mathbf{X}_n$ close to zero frequency because the effects of differenced measurement error noises $(\mathbf{v}_i - \mathbf{v}_{i-1})$ are negligible at zero frequency. This method differs from some existing procedures that consider the decomposition of time series only in the time domain. Our arguments can be justified by using the frequency decomposition of $\mathbf{y}_i$ and $\mathbf{r}_i^{(n)} = \Delta \mathbf{y}_i$ $(= \mathbf{y}_i - \mathbf{y}_{i-1}$ and $\mathbf{y}_0$ being fixed). We discuss this issue in Section 4 (refer to Section 5.2 of Kunitomo and Sato (2021)).

3.2 Backward Smoothing

We investigate the role of the initial condition in the non-stationary process and consider the situation when the time is reversed, that is, from n to 0, rather than from 0 to n. We take the $n \times p$ matrix $\mathbf{Y}_n^* = (\mathbf{y}_{i-1}^{'})$ and set the $np \times 1$ random vector $(\mathbf{y}_0^{'}, \cdots, \mathbf{y}_{n-1}^{'})^{'}$ [1]. Given the initial condition $\mathbf{y}_n$, we rewrite

$$\text{(3.5)} \qquad \mathrm{vec}(\mathbf{Y}_n^*) \sim N_{n \times p}\left(\mathbf{1}_n \cdot \mathbf{y}_n^{'}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}_v + \mathbf{C}_n^{'}\mathbf{C}_n \otimes \boldsymbol{\Sigma}_x\right) ,$$

where $\mathbf{1}_n^{'} = (1, \cdots, 1)$ and $\mathbf{C}_n$ is given by (2.4).
We use $\mathbf{K}_n^*$−transformation that from $\mathbf{Y}_n^*$ to $\mathbf{Z}_n^*$ $(= (\mathbf{z}_k^{*'}))$ by

$$\text{(3.6)} \qquad \mathbf{Z}_n^* = \mathbf{K}_n^*\left(\mathbf{Y}_n^* - \bar{\mathbf{Y}}_n^*\right) , \mathbf{K}_n^* = \mathbf{P}_n^*\mathbf{C}_n^{'-1} ,$$

where $\bar{\mathbf{Y}}_n^* = \mathbf{1}_n \mathbf{y}_n^{'}$,

$$\text{(3.7)} \qquad \mathbf{C}_n^{'-1} = \begin{pmatrix} 1 & -1 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}_{n \times n} ,$$

and the (k, j)−th element of $\mathbf{P}_n^* = (p_{kj}^{(n)})$ is defined by

$$\text{(3.8)} \qquad p_{kj}^{*(n)} = \sqrt{\frac{2}{n + \frac{1}{2}}} \sin\left[\frac{2\pi}{2n+1}(k - \frac{1}{2})j\right] .$$

By using the spectral decomposition, $\mathbf{C}_n^{'-1}\mathbf{C}_n^{-1} = \mathbf{P}_n^{*'}\mathbf{D}_n\mathbf{P}_n^*$, and $\mathbf{D}_n$ is a diagonal matrix with the k-th element $d_k = 2[1 - \cos(\pi(\frac{2k-1}{2n+1}))]$ $(k = 1, \cdots, n)$. Then, we

[1] Given the initial condition $\mathbf{y}_n$, we consider the joint distribution of $(\mathbf{y}_{n-1}^{'}, \cdots, \mathbf{y}_0^{'})^{'}$, while we took $\mathbf{y}_i$ $(i = 0, 1, \cdots, n)$ in Section 3.1.

write

$$a^*_{kn}\ (= d_k)\ = 4\sin^2\left[\frac{\pi}{2}\left(\frac{2k-1}{2n+1}\right)\right]\ (k=1,\cdots,n)\ . \tag{3.9}$$

(Refer to Appendix for the derivation.)

We consider the partial inversion of the transformed orthogonal processes. Let an $n \times p$ matrix be

$$\hat{\mathbf{X}}^*_n(\mathrm{Q}_n) = \mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{Q}_n\mathbf{P}_n^*\mathbf{C}_n^{'-1}(\mathbf{Y}^*_n - \bar{\mathbf{Y}}^*_n) \tag{3.10}$$

and

$$\mathbf{Z}^*_n = \mathbf{P}^*_n\mathbf{C}_n^{'-1}(\mathbf{Y}^*_n - \bar{\mathbf{Y}}^*_n)\ , \mathbf{Y}^*_n = \bar{\mathbf{Y}}^*_n + \mathbf{X}^*_n + \mathbf{V}^*_n\ , \tag{3.11}$$

where $\mathbf{X}^*_n = (\mathbf{x}^{*'}_{i-1})$ and $\mathbf{V}^*_n = (\mathbf{v}^{*'}_{i-1})$ are the $n\times p$ matrices, and $\mathbf{x}^*_{i-1} = \mathbf{x}_{i-1} - \mathbf{x}_n$ $(i = 1,\cdots,n)$.

The stochastic process $\mathbf{Z}^*_n$ is the orthogonal decomposition of the original time series $\mathbf{Y}^*_n$ in the frequency domain, and $\mathbf{Q}_n$ is an $n \times n$ filtering matrix. Because $\mathbf{Y}^*_n$ consists of non-stationary time series, we need a special form of transformation $\mathbf{K}^*_n$. We provide an explicit form for the trend filtering procedure. Then, let the $n \times p$ matrix be

$$\hat{\mathbf{X}}^*_n(m) = \mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{Q}_n^{(m)}\mathbf{P}_n^*\mathbf{C}_n^{'-1}(\mathbf{Y}^*_n - \bar{\mathbf{Y}}^*_n) \tag{3.12}$$

and $\mathbf{Q}_n^{(m)} = \mathbf{J}_m^{'}\mathbf{J}_m$.

We construct an estimator of $n \times p$ hidden state matrix $\mathbf{X}^*_n$ only in the lower-frequency parts by using the inverse transformation of $\mathbf{Z}^*_n$ and by deleting the estimated noise parts. We denote the hidden trend state as

$$\mathbf{X}^*_n(m) = \mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{Q}_n^{(m)}\mathbf{P}_n^*\mathbf{C}_n^{'-1}\mathbf{X}^*_n\ . \tag{3.13}$$

3.3 Initial Value Problem and Convergence

When we have non-stationary time series observation that follows a random walk as a statistical model, the role of the initial value is important because of non-stationarity. This aspect is different from stationary time series models, in which the effects of the initial value are negligible when the sample size is large. Hence, a smoothing or filtering procedure of non-stationary time series, that does not depend much on the initial value, is important. As the initial value, there can be two possibilities as $\mathbf{y}_0$ and $\mathbf{y}_n$ when we have $n+1$ vector observations $\mathbf{y}_i$ $(i = 0,1,\cdots,n)$. In this problem, we have an interesting useful result.

We consider two operators $T_{2k}^{(m,n)}$ and $T_{2k-1}^{(m,n)} (k \geq 1)$ to an $n \times 1$ vector. Let $T_0 = I_n$ and define $T_{2k-1}^{(m,n)}$ and $T_{2k}^{(m,n)}$ recursively for $k = 1,\cdots,M$ by

$$T_{2k+1}^{(m,n)}(\mathbf{y}) = \mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n^{(m)}\mathbf{P}_n\mathbf{C}^{-1}[\mathbf{y} - \mathbf{1}_n(\mathbf{e}_1^{'}T_{2k}^{(m,n)}(\mathbf{y}))] + \mathbf{1}_n(\mathbf{e}_1^{'}T_{2k}^{(m,n)}(\mathbf{y}))\ , \tag{3.14}$$

and

$$(3.15)\, T_{2k}^{(m,n)}(\mathbf{y}) = \mathbf{C}_n^{'}\mathbf{P}_n^{*}{}^{'}\mathbf{Q}_n^{(m)}\mathbf{P}_n^{*}\mathbf{C}^{'-1}[\mathbf{y}-\mathbf{1}_n(\mathbf{e}_n^{'}T_{2k-1}^{(m,n)}(\mathbf{y}))] + \mathbf{1}_n(\mathbf{e}_n^{'}T_{2k-1}^{(m,n)}(\mathbf{y}))\ ,$$

where $\mathbf{Q}_n^{(m)} = \mathbf{J}_m^{'}\mathbf{J}_m$, $\mathbf{1}_n^{'} = (1,\cdots,1)$, and $\mathbf{e}_1^{'} = (1,0,\cdots,0)$ and $\mathbf{e}_n^{'} = (0,\cdots,0,1)$ are unit vectors.
The operator $T_{2k-1}^{(m,n)}$ $(k \geq 1)$ is the SIML filtering with the initial value at $i = 0$ $\mathbf{y}_0$ and $T_{2k}^{(m,m)}$ $(k \geq 1)$ is the backward filtering with the initial value at $i = n$. For non-stationary time series, two operators have different roles in smoothing procedure. We can repeat the smoothing procedures such that for $k \geq 1$

$$\begin{aligned} T_{2k+1}^{(m,n)}(\mathbf{y}) &= \mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n^{(m)}\mathbf{P}_n\mathbf{C}^{-1}\mathbf{y} + [\mathbf{I}_n - \mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n^{(m)}\mathbf{P}_n\mathbf{C}^{-1}] \\ &\quad \times\mathbf{1}_n\mathbf{e}_1^{'}\left[\mathbf{C}_n^{'}\mathbf{P}_n^{*}{}^{'}\mathbf{Q}_n^{(m)}\mathbf{P}_n^{*}\mathbf{C}^{'-1}\left(\mathbf{y}-\mathbf{1}_n(\mathbf{e}_n^{'}T_{2k-1}^{(m,n)}(\mathbf{y}))\right) + \mathbf{1}_n(\mathbf{e}_n^{'}T_{2k-1}^{(m,n)}(\mathbf{y}))\right]\ . \end{aligned}$$

Then, we have the next proposition on the convergence of the smoothing procedure and the proof is given in the Appendix.

Theorem 3.1 : As $k \to \infty$, there exists n_0 such that, for $n_0 < n$ and $m < n$, we have

$$(3.16) \qquad T_{2k+1}^{(m,n)} \to T_{1*}^{(m,n)} = \sum_{s=0}^{\infty}(\mathbf{A}_2^{(m,n)})^s\mathbf{A}_1^{(m,n)}\ ,$$

and

$$(3.17) \qquad T_{2k}^{(m,n)} \to T_{2*}^{(m,n)} = \sum_{s=0}^{\infty}(\mathbf{A}_{2*}^{(m,n)})^s\mathbf{A}_{1*}^{(m,n)}\ ,$$

where

$$\begin{aligned} \mathbf{A}_1^{(m,n)} &= \mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n^{(m)}\mathbf{P}_n\mathbf{C}^{-1} + [\mathbf{I}_n - \mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n^{(m)}\mathbf{P}_n\mathbf{C}^{-1}]\mathbf{1}_n\mathbf{e}_1^{'}\mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{Q}_n^{(m)}\mathbf{P}_n^{*}\mathbf{C}^{'-1}\ , \\ \mathbf{A}_2^{(m,n)} &= [\mathbf{I}_n - \mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n^{(m)}\mathbf{P}_n\mathbf{C}^{-1}]\mathbf{1}_n \times [1-\mathbf{e}_1^{'}\mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{Q}_n^{(m)}\mathbf{P}_n^{*}\mathbf{C}^{'-1}\mathbf{1}_n]\mathbf{e}_n^{'}\ , \\ \mathbf{A}_{1*}^{(m,n)} &= \mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{Q}_n^{(m)}\mathbf{P}_n^{*}\mathbf{C}^{'-1} + [\mathbf{I}_n - \mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{Q}_n^{(m)}\mathbf{P}_n^{*}\mathbf{C}^{'-1}]\mathbf{1}_n\mathbf{e}_n^{'}\mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n^{(m)}\mathbf{P}_n\mathbf{C}^{-1}\ , \\ \mathbf{A}_{2*}^{(m,n)} &= [\mathbf{I}_n - \mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{Q}_n^{(m)}\mathbf{P}_n^{*}\mathbf{C}^{'-1}]\mathbf{1}_n \times [1-\mathbf{e}_n^{'}\mathbf{C}_n\mathbf{P}_n\mathbf{Q}_n^{(m)}\mathbf{P}_n\mathbf{C}^{-1}\mathbf{1}_n]\mathbf{e}_n^{'}\ . \end{aligned}$$

The absolute values of all eigenvalues of $\mathbf{A}_2^{(m,n)}$ and $\mathbf{A}_{2*}^{(m,n)}$ are less than one. Then, we can express

$$\sum_{s=0}^{\infty}(\mathbf{A}_2^{(m,n)})^s = (\mathbf{I}_n - \mathbf{A}_2^{(m,n)})^{-1}\ ,\ \sum_{s=0}^{\infty}(\mathbf{A}_{2*}^{(m,n)})^s = (\mathbf{I}_n - \mathbf{A}_{2*}^{(m,n)})^{-1}\ .$$

Given that the initial value is the starting point of non-stationary time series, we need to develop a smoothing procedure that does not depend on the initial value. Practically, often we do want to use the procedure that does not depend on the first or latest observation $\mathbf{y}_0$ or $\mathbf{y}_n$. In these cases, it may be reasonable to use the $T_2^{(n)}$

or $T_1^{(n)}$, respectively.

It may be interesting to find the difference between the two procedures of smoothing. Let the two operators be $\mathbf{H}_n = (h_{ab}^{(n)}) = \mathbf{P}_n \mathbf{Q}_n^{(m)} \mathbf{P}_n$ and $\mathbf{F}_n = (f_{ab}^{(n)}) = \mathbf{P}_n^{*'} \mathbf{Q}_n^{(m)} \mathbf{P}_n^{*}$. Then, each term of $h_{ab}^{(n)}$ and $f_{ab}^{(n)}$ are sums of m trigonometric functions in the forward and backward SIML smoothing. They are similar as we summarize in the next result. The proof is given in the Appendix.

Theorem 3.2 : (i) Assume that $n/m_n \to \infty$ as $n \to \infty$. Then,

$$(3.18) \qquad (\frac{n}{m_n}) \max_{a,b} |h_{ab}^{(n)} - f_{ab}^{(n)}| \to 0 \ .$$

(ii) Define the forward and backward operators by $\mathbf{H}_n^{(0)} = \mathbf{C}_n \mathbf{H}_n \mathbf{C}_n^{-1}$ and $\mathbf{F}_n^{*} = \mathbf{C}_n^{'} \mathbf{F}_n \mathbf{C}_n^{-1'}$, respectively. For any $\delta > 0$, we take $m = m_n$ such that $0 < m_n < m_n^{1+\delta} < n$. Then, as $n \to \infty$,

$$(3.19) \qquad (\frac{n}{m_n^{1+\delta}}) \mathrm{Tr}[\mathbf{H}_n^{(0)} - \mathbf{F}_n^{*}] \to 0 \ ,$$

where the trace of an $n \times n$ matrix $\mathbf{A} = (a_{ij})$ is defined by $\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$.

From this result, the backward SIML smoothing is essentially the same method as the forward smoothing. As we shall see in Section 4.2, it is a real (finite- and discrete) Fourier transformation if we take that the time is reversed from n to 0, rather than from 0 to n.

3.4 Band Smoothing

We consider a general filtering based on the $\mathbf{K}_n$ and $\mathbf{K}_n^{*}$ transformations and use the inversion of some frequency parts of the random matrix $\mathbf{Z}_n$ and $\mathbf{Z}_n^{*}$. The leading example is the seasonal frequency in the discrete time series, and we take s (> 1) as a positive integer.
Let an $m_2 \times [m_1 + m_2 + (n - m_1 - m_2)]$ choice matrix be $\mathbf{J}_{m_1,m_2} = (\mathbf{O}, \mathbf{I}_{m_2}, \mathbf{O})$ (we take $m_1 + m_2 < n$), and let also $n \times p$ matrices be

$$(3.20) \qquad \hat{\mathbf{X}}_n(m_1, m_2) = \mathbf{C}_n \mathbf{P}_n \mathbf{J}_{m_1,m_2}^{'} \mathbf{J}_{m_1,m_2} \mathbf{P}_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0)$$

and

$$(3.21) \qquad \hat{\mathbf{X}}_n^{*}(m_1, m_2) = \mathbf{C}_n^{'} \mathbf{P}_n^{*'} \mathbf{J}_{m_1,m_2}^{'} \mathbf{J}_{m_1,m_2} \mathbf{P}_n^{*} \mathbf{C}_n^{'-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_n)$$

and an $n \times n$ matrix $\mathbf{Q}_n = \mathbf{Q}_n^{(m_1,m_2)} = \mathbf{J}_{m_1,m_2}^{'} \mathbf{J}_{m_1,m_2}$.

As an example in economic data, when we have the seasonal frequency s (> 1), we can take $m_1 = [2n/s] - [m/2]$ and $m_2 = m$. For instance, we take $s = 4$ for the quarterly data and $s = 12$ for the monthly data. (See Sato and Kunitomo (2021).) Similar to the trend smoothing problem, the SIML-filtering value

$$\mathbf{X}_n(m_1, m_2) = \mathbf{C}_n \mathbf{P}_n \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n \mathbf{C}_n^{-1} \mathbf{X}_n^{(0)} , \tag{3.22}$$

and

$$\mathbf{X}_n^*(m_1, m_2) = \mathbf{C}_n^{'} \mathbf{P}_n^{*'} \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n^* \mathbf{C}_n^{'-1} \mathbf{X}_n^* , \tag{3.23}$$

respectively, which are based on the estimated frequency components of $\mathbf{x}_i^{(0)}$ $(i = 1, \cdots, n)$ or $\mathbf{x}_i^*$ $(i = 0, \cdots, n-1)$.
In this case, we can define T_{2k-1} and T_{2k} for $k = 1, \cdots M$ as (3.14) and (3.15) by using $\mathbf{J}_{m_1,m_2}$, rather than $\mathbf{J}_m$. Then, it is straightforward to to find the next proposition on the convergence of smoothing procedure, and the proof is in the Appendix.

Theorem 3.3 : As $k \to \infty$, there exists n_0 such that, for $n_0 < n$, we have

$$T_{2k+1}^{(m_1,m_2,n)} \to T_{1*}^{(m_1,m_2,n)} = \sum_{s=0}^{\infty} (\mathbf{A}_2^{(m_1,m_2,n)})^s \mathbf{A}_1^{(m_1,m_2,n)} , \tag{3.24}$$

and

$$T_{2k}^{(m_1,m_2,n)} \to T_{2*}^{(m_1,m_2,n)} = \sum_{s=0}^{\infty} (\mathbf{A}_{2*}^{(m_1,m_2,n)})^s \mathbf{A}_{1*}^{(m_1,m_2,n)} , \tag{3.25}$$

where

$$\begin{aligned}
\mathbf{A}_1^{(m_1,m_2,n)} &= \mathbf{C}_n \mathbf{P}_n \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n \mathbf{C}^{-1} \\
&+ [\mathbf{I}_n - \mathbf{C}_n \mathbf{P}_n \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n \mathbf{C}^{-1}] \mathbf{1}_n \mathbf{e}_1^{'} \mathbf{C}_n^{'} \mathbf{P}_n^* \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n^* \mathbf{C}^{'-1} , \\
\mathbf{A}_2^{(m_1,m_2,n)} &= [\mathbf{I}_n - \mathbf{C}_n \mathbf{P}_n \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n \mathbf{C}^{-1}] \mathbf{1}_n \times [1 - \mathbf{e}_1^{'} \mathbf{C}_n^{'} \mathbf{P}_n^{*'} \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n^* \mathbf{C}^{'-1} \mathbf{1}_n] \mathbf{e}_n^{'} ,
\end{aligned}$$

$$\begin{aligned}
\mathbf{A}_{1*}^{(m_1,m_2,n)} &= \mathbf{C}_n^{'} \mathbf{P}_n^{*'} \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n^* \mathbf{C}^{'-1} \\
&+ [\mathbf{I}_n - \mathbf{C}_n^{'} \mathbf{P}_n^{*'} \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n^* \mathbf{C}^{'-1}] \mathbf{1}_n \mathbf{e}_n^{'} \mathbf{C}_n \mathbf{P}_n \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n \mathbf{C}^{-1} \\
\mathbf{A}_{2*}^{(m_1,m_2,,n)} &= [\mathbf{I}_n - \mathbf{C}_n^{'} \mathbf{P}_n^{*'} \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n^* \mathbf{C}^{'-1}] \mathbf{1}_n \times [1 - \mathbf{e}_n^{'} \mathbf{C}_n \mathbf{P}_n \mathbf{Q}_n^{(m_1,m_2)} \mathbf{P}_n \mathbf{C}^{-1} \mathbf{1}_n] \mathbf{e}_n^{'} .
\end{aligned}$$

The absolute values of all eigenvalues of $\mathbf{A}_2^{(m_1,m_2,n)}$ and $\mathbf{A}_{2*}^{(m_1,m_2,n)}$ are less than one; then, we can express

$$\sum_{s=0}^{\infty} (\mathbf{A}_2^{(m_1,m_2,n)})^s = (\mathbf{I}_n - \mathbf{A}_2^{(m_1,m_2,n)})^{-1} , \sum_{s=0}^{\infty} (\mathbf{A}_{2*}^{(m_1,m_2,n)})^s = (\mathbf{I}_n - \mathbf{A}_{2*}^{(m_1,m_2,n)})^{-1} .$$

This result would be useful for handling seasonality of economic time series, as an example. Theorem 3.1 can be regarded as a special case of Theorem 3.3 when $m_1 = 0$ and $m_2 = m$.

3.5 Multi-step Smoothing

In the forward and backward smoothing procedures, choosing an appropriate m is important. Howver, the problem becomes difficult when seasonal components exist. Then, it may be normal to repeat smoothing because the forward and backward smoothing several times, which may be called *multi-stage smoothing*, can be run.

Let $T_{1*}^{(m_1,n)}$ be the first stage forward smoothing with a specific choice of m_1. Then, we can define the double-stage forward smoothing by

$$T_{1,1}^{(m_1,m_2,n)} = T_{1*}^{(m_1,n)} T_{1*}^{(m_2,n)} . \tag{3.26}$$

Similarly, we can define the double-stage backward smoothing by

$$T_{2,2}^{(m_1,m_2,n)} = T_{2*}^{(m_1,n)} T_{2*}^{(m_2,n)} . \tag{3.27}$$

More complicated smoothing procedures can also exist. Hence, we need some criterion to find an appropriate smoothing procedure for applications. Consequently, handling complicated seasonal patterns in the frequency domain is possible, for instance, because we first take a rather large m_1 and then we take a smaller m_2.

For real applications, finding an appropriate m or m_1 and m_2 at the beginning might not be certain. Section 6.1 of Kunitomo and Sato (2021) has discussed a guide on choosing frequencies. In addition, at the first stage, one strategy in the trend estimation would be to choose a relatively large m_1, which should be less than the seasonality frequency. Then, at the second stage, we choose m_2, which is smaller than m_1 and use the following evaluation criterion.

3.6 Prediction Errors and Evaluation Criteria

The problem of choosing an appropriate filtering, including the choice of m (or m_1 and m_2 in a more general case) in smoothing, is an important question for applications. Given that our procedure does not assume a particular distribution such as Gaussianity and semi-parametric, it finds a challenging one. As we discuss in the next section (refer to Kunitomo and Sato (2021)), a typical method to handle the problem exists, which has been illustrated by the forward filtering case. However, in the backward smoothing case, we have a similar argument.

Let $\mathbf{r}_j^{(n)} = \mathbf{y}_j^{(n)} - \mathbf{y}_{j-1}^{(n)}$ $(j = 1, \cdots, n)$; hence we write

$$\hat{\mathbf{r}}_j^{(n)} = \sum_{k=1}^{n} p_{jk} \mathbf{z}_k , \tag{3.28}$$

where $\mathbf{z}_k^*$ is the orthogonal process at the frequency $\lambda_k^{(n)} = (k - 1/2)/(2n+1)$ $(k = 1, \cdots, n)$ (refer to Section 5 of Kunitomo and Sato (2021)).

Then, for $h \geq 1$, it may be standard to use the predictor

$$\hat{\mathbf{r}}_{n+h}^{(n)}(m) = \sum_{k=1}^{m} p_{n+h,k}\mathbf{z}_k \ , \tag{3.29}$$

which is a linear combination of m orthogonal processes with different frequencies. Then, for $h \geq 1$, it may be reasonable to use the linear predictor

$$\hat{\mathbf{x}}_{n+h}^{(n)}(m) = \sum_{s=h+1}^{n+h} \hat{\mathbf{r}}_s^{(n)}(m) = \sum_{s=h+1}^{n+h} \sum_{k=1}^{m} p_{sk}\mathbf{z}_k \ . \tag{3.30}$$

By using (3.1) and (3.2), the prediction error can be written as

$$\hat{\mathbf{x}}_{n+h}^{(n)}(m) - \mathbf{x}_{n+h}^{(n)} = \sum_{k=1}^{m} \sum_{s=h+1}^{n+h} \sum_{j=1}^{n} p_{sj}(\mathbf{C}_n^{-1}\mathbf{V}_n)_{kj} \ . + \sum_{k=m+1}^{n} \sum_{s=h+1}^{n+h} \sum_{j=1}^{n} p_{sj}(\mathbf{C}_n^{-1}\mathbf{X}_n)_{kj} \ .$$

We use an elementary relation that

$$\sum_{s=h+1}^{n+h} p_{sk} = \frac{1}{\sqrt{2n+1}} \frac{\sin \frac{2\pi}{2n+1}(n+h)(k-\frac{1}{2}) - \sin \frac{2\pi}{2n+1}h(k-\frac{1}{2})}{\sin \frac{2\pi}{2n+1}\frac{1}{2}(k-\frac{1}{2})} \ .$$

Then, when $p = 1$, for example, by using a_{kn}^* $(k = 1, \cdots, m)$ in (2.8), we can derive the prediction MSE as

$$\begin{aligned} MSE(m) &= \frac{4\sigma_v^2}{2n+1} \sum_{k=1}^{m} [\sin \frac{2\pi}{2n+1}(n+h)(k-\frac{1}{2}) - \sin \frac{2\pi}{2n+1}h(k-\frac{1}{2})]^2 \\ &\quad + \frac{\sigma_x^2}{2n+1} \sum_{k=m+1}^{n} [\frac{\sin \frac{2\pi}{2n+1}(n+h)(k-\frac{1}{2}) - \sin \frac{2\pi}{2n+1}h(k-\frac{1}{2})}{\sin \frac{2\pi}{2n+1}\frac{1}{2}(k-\frac{1}{2})}]^2 \ . \end{aligned} \tag{3.31}$$

As a typical example, we set $\sigma_v^2 = 2, \sigma_x^2 = 1, h = 4, n = 100$. The minimum value of MSE is attained when $m^* = 23$.
We notice that the first term is an increasing function of m, while the second term is a decreasing function of m. A point of m^* exists such that $MSE(m)$ is minimized. Several criteria that are based on the prediction MSE exist. Because the prediction error depends on the unknown parameters of $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_v$, we must replace them in a simple manner. When $p = 1$, we need the ratio of estimated variances, which were constructed in the discussion of Section 3 of Kunitomo, Awaya, and Kurisu (2019).

4. Discussions

4.1 Extended Errors-in-Variables Models

Possible generalizations of the basic model in Section 3 exist. In this section, we consider the additive decomposition model

$$\mathbf{y}_i = \mathbf{x}_i + \mathbf{s}_i + \mathbf{v}_i \quad (i = 0, 1, \cdots, n), \tag{4.1}$$

where $\Delta \mathbf{x}_i = \mathbf{v}_i^{(x)}$ and we take positive integers s $(s > 1)$, N, and $n = sN$ for the resulting simplicity. Furthermore, we consider that the noise component $\mathbf{v}_i$ $(i = 0, 1, \cdots, n)$ and seasonal component $\mathbf{s}_i$ $(i = 0, 1, \cdots, n)$ are sequences of the stationary processes, which satisfy

$$\mathbf{v}_i = \sum_{j=-\infty}^{\infty} \mathbf{C}_j^{(v)} \mathbf{e}_{i-j}^{(v)} , \tag{4.2}$$

and

$$\mathbf{s}_i = \sum_{j=-\infty}^{\infty} \mathbf{C}_{sj}^{(s)} \mathbf{e}_{i-sj}^{(s)} , \tag{4.3}$$

where $\mathbf{e}_i^{(v)}$ and $\mathbf{e}_i^{(s)}$ are sequences of i.i.d. random vectors with $\mathbf{E}(\mathbf{e}_i^{(v)}) = \mathbf{E}(\mathbf{e}_i^{(s)}) = \mathbf{0}$, and $\mathbf{E}(\mathbf{e}_i^{(v)} \mathbf{e}_i^{(v)'}) = \mathbf{\Sigma}_e^{(v)}$ (a positive definite matrix), and $\mathbf{E}(\mathbf{e}_i^{(s)} \mathbf{e}_i^{(s)'}) = \mathbf{\Sigma}_e^{(s)}$ (a non-negative definite matrix).
The $p \times p$ coefficient matrices $\mathbf{C}_j^{(v)}$ and $\mathbf{C}_j^{(s)}$ are absolutely summable such that $\sum_{j=-\infty}^{\infty} \|\mathbf{C}_j^{(v)}\| < \infty$, and $\sum_{j=-\infty}^{\infty} \|\mathbf{C}_{sj}^{(s)}\| < \infty$, where $\|\mathbf{C}_j^{(v)}\| = \max_{k,l=1,\cdots,p} |c_{k,l}^{(v)}(j)|$ for $\mathbf{C}_j^{(v)} = (c_{k,l}^{(v)}(j))$, and $\|\mathbf{C}_{sj}^{(s)}\| = \max_{k,l=1,\cdots,p} |c_{k,l}^{(s)}(j)|$ for $\mathbf{C}_{sj}^{(s)} = (c_{k,l}^{(s)}(sj)))$, respectively.

Let $\mathbf{f}_{\Delta x}(\lambda)$, $\mathbf{f}_s(\lambda)$, and $\mathbf{f}_v(\lambda)$ be the spectral density $(p \times p)$ matrices of $\Delta \mathbf{x}_i$, $\mathbf{s}_i$ and $\mathbf{v}_i$ $(i = 1, \cdots, n)$. Then

$$\mathbf{f}_v(\lambda) = (\sum_{j=-\infty}^{\infty} \mathbf{C}_j^{(s)} e^{2\pi i \lambda j}) \mathbf{\Sigma}_e^{(v)} (\sum_{j=-\infty}^{\infty} \mathbf{C}_j^{(v)'} e^{-2\pi i \lambda j}) \quad (-\frac{1}{2} \le \lambda \le \frac{1}{2}) , \tag{4.4}$$

and

$$\mathbf{f}_s(\lambda) = (\sum_{j=-\infty}^{\infty} \mathbf{C}_{sj}^{(s)} e^{2\pi i \lambda sj}) \mathbf{\Sigma}_e^{(s)} (\sum_{j=-\infty}^{\infty} \mathbf{C}_{sj}^{(s)'} e^{-2\pi i \lambda sj}) \quad (-\frac{1}{2} \le \lambda \le \frac{1}{2}) , \tag{4.5}$$

where we set $\mathbf{C}_0^{(v)} = \mathbf{C}_0^{(s)} = \mathbf{I}_p$ as normalizations and $i^2 = -1$.
Then, the relation between the $p \times p$ spectral density matrix of the transformed vector process, which are observable, and the spectral density of the observed difference series $\Delta \mathbf{y}_i$ $(= \mathbf{y}_i - \mathbf{y}_{i-1})$ can be represented as

$$\mathbf{f}_{\Delta y}(\lambda) = \mathbf{f}_{\Delta x}(\lambda) + (1 - e^{2\pi i \lambda})[\mathbf{f}_s(\lambda) + f_v(\lambda)](1 - e^{-2\pi i \lambda}) \ . \tag{4.6}$$

We denote the long-run variance-covariance matrices of trend components and stationary components for $g, h = 1, \cdots, p$ as

$$\text{(4.7)} \qquad \mathbf{\Sigma}_x = \mathbf{f}_{\Delta x}(0) \; (= (\sigma_{gh}^{(x)})) \; , \; \mathbf{\Sigma}_v = f_v(0) \; = (\sigma_{gh}^{(v)}) \; .$$

Let $f_v^{(SR)}(\lambda_k)$, $f_s^{(SR)}(\lambda_k)$, and $f_{\Delta x}^{(SR)}(\lambda_k)$ be the symmetrized $p \times p$ spectral matrices of $\mathbf{v}_i$, $\mathbf{s}_i$ and $\Delta \mathbf{x}_i$ at λ_k $(= (k - \frac{1}{2})/(2n+1))$ for $k = 1, \cdots, n$, that is, $f_v^{(SR)}(\lambda_k) = (1/2)[f_v^{(SR)}(\lambda_k) + \bar{f}_v^{(SR)}(\lambda_k)]$, $f_s^{(SR)}(\lambda_k) = (1/2)[f_s^{(SR)}(\lambda_k) + \bar{f}_s^{(SR)}(\lambda_k)]$ and $f_{\Delta x}^{(SR)}(\lambda_k) = (1/2)[f_{\Delta x}^{(SR)}(\lambda_k) + \bar{f}_{\Delta x}^{(SR)}(\lambda_k)]$.
Theorem 5.1 of Kunitomo and Sato (2021) offers the condition that the orthogonal processes are approximately distributed as the Gaussian distribution. Then, (-2) times the log-likelihood function in the general model can be approximated as

$$\text{(4.8)} \quad (-2)l_n(\boldsymbol{\theta}) \;=\; \sum_{k=1}^{n} \log |a_{kn}^{*}(f_v^{(SR)}(\lambda_k) + f_s^{(SR)}(\lambda_k)) + f_{\Delta x}^{(SR)}(\lambda_k)| + \sum_{k=1}^{n} \mathbf{z}_k^{'}[a_{kn}^{*}(f_v^{(SR)}(\lambda_k) + f_s^{(SR)}(\lambda_k)) + f_{\Delta x}^{(SR)}(\lambda_k)]^{-1}\mathbf{z}_k \; .$$

In the forward smoothing, Kunitomo and Sato (2021) have used the causal MA representation of the stationary process and discussed its interpretation in their Section 5. In the backward smoothing, we need the non-causal MA representation of the stationary process. For causal and non-causal MA models, we refer to Chapter 4 of Brockwell and Davis (1990). However, we have a similar interpretation of the backward smoothing based on the frequency domain analysis, which shall be discussed in the next subsection.

4.2 Frequency Interpretation

The SIML smoothing method might be regarded initially as an *ad-hoc* statistical procedure without any mathematical foundation. However, conversely, there is a rather solid statistical foundation exists. Section 5 of Kunitomo and Sato (2021a) has discussed the justification of the SIML forward-smoothing, and it may differ from the standard usage of traditional time series analysis in the frequency domain. (Doob (1953), and Brockwell and Davis (1990), and its extensions to the non-stationary process, and Brillinger and Hatanaka (1969), Brillinger (1980) for related topics.)

We can proceed a similar argument as Kunitomo and Sato (2021), but it is on the backward smoothing. For $\lambda_k^{(n)} = (k - 1/2)/(2n+1)$ $(k = 1, \cdots, n)$, we write

$$\text{(4.9)} \qquad \mathbf{z}_n^{*}(\lambda_k^{(n)}) = \sum_{j=1}^{n} \mathbf{r}_{j-1}^{(n)*}[\frac{2}{\sqrt{2n+1}} \sin[2\pi\lambda_k^{(n)} j] \;\; (k = 1, \cdots, n),$$

where $\mathbf{r}_{j-1}^{(n)*} = \mathbf{y}_{j-1}^{(n)} - \mathbf{y}_j^{(n)}$ $(j = 1, \cdots, n)$.
(We note that under the assumption of (4.1)-(4.3), $\mathbf{r}_j^{(n)*}$ is a stationary process and

has a MA representation.)
Then, by using the inversion transformation with $\mathbf{P}_n^*$, we can confirm that

$$\mathbf{r}_{s-1}^{(n)*} = \sum_{k=1}^{n} p_{sk}^* \mathbf{z}_n^* (\lambda_k^{(n)}) \;\; (s = 1, \cdots, n). \tag{4.10}$$

It is another representation of $\mathbf{R}_n^* = (\mathbf{r}_{i-1}^{*(n)'}) = \mathbf{C}_n^{'-1} \hat{\mathbf{X}}_n^* (\mathrm{Q}^*)$ when $\mathbf{Q}_n^* = \mathbf{I}_n$. For any $s-1$ $(s = 1, \cdots, n)$, $\mathbf{r}_{s-1}^{(n)}*$ can be recovered as the weighted sum of orthogonal processes $\mathbf{z}^{*(n)}(\lambda_k^{(n)})$ at frequency $\lambda_k^{(n)}$ $(k = 1, \cdots, n)$.
Then, by using $\mathbf{Y}_n^* = \mathbf{C}_n^{'} \mathbf{R}_n^{'}$, we recover the non-stationary process $\mathbf{y}_t^{(n)}$ $(t = 0, \cdots, n-1)$ given the initial condition $\mathbf{y}_0$ as

$$\mathbf{y}_t^{(n)} = \mathbf{y}_n + \sum_{s=1}^{t} \mathbf{r}_{n-s}^{*(n)} \;\; . \tag{4.11}$$

Let

$$\alpha_n(\lambda_m^{(n)}, j) = \frac{1}{n} \sum_{k=1}^{m} [2 \sin 2\pi \lambda_k^{(n)} j] \tag{4.12}$$

Then, when $\lambda_m^{(n)} \to \lambda$ as $n \to \infty$ $(0 < \lambda < \frac{1}{2})$, we find

$$\beta_n(\lambda_m^{(n)}, j) \to \beta(\lambda, j) = \frac{2[1 - \cos 2\pi\lambda \; j]}{\pi \; j} \; .$$

If we set the uncorrelated stochastic process of uncorrelated increments with continuous parameter λ $(0 \leq \lambda \leq \frac{1}{2})$ as $B_n(\lambda) = \sum_{j=1}^{n} \beta(\lambda, j) \mathbf{r}_{j-1}^{*(n)}$, then we find

$$\int_0^{\frac{1}{2}} \sin[2\pi\lambda s] dB_n(\lambda) = \mathbf{r}_{s-1}^{*(n)} \;\; (s = 1, \cdots, n) \; . \tag{4.13}$$

This corresponds to the continuous representation of a discrete (real-valued) stationary time series in the frequency domain (refer to Chapter 7.4 of Anderson (1971)). If we write the limit of $\mathbf{B} = \lim_{n \to \infty} \mathbf{B}_n(\lambda)$ (assuming it exists), then the (real-valued) spectral distribution matrix F_{RS} for any $0 \leq \lambda_1 < \lambda_2 \leq 1/2$ can be defined as

$$F_{RS}(\lambda_2 - \lambda_1) = \mathbf{E}[(\mathbf{B}(\lambda_2 - \lambda_1) \mathbf{B}(\lambda_2 - \lambda_1)^{'}] = \int_{\lambda_1}^{\lambda_2} f_{RS}(\lambda) d\lambda \tag{4.14}$$

if F_{RS} is absolutely continuous and the matrix-valued density process $f_{RS}(\lambda)$ $(0 \leq \lambda_1 < \lambda_2 \leq 1/2)$ exists.
We set $\hat{\mathbf{R}}_n^*(m) = (\hat{\mathbf{r}}_i^{*(m,n)'}) = \mathbf{C}_n^{'-1} \hat{\mathbf{X}}_n^*(m)$, where $\mathbf{r}_i^{*(m,n)}$ are $p \times 1$ vectors for $i = 1, \cdots, n$. If we write

$$\hat{\mathbf{r}}_{s-1}^{*(m,n)} = \sum_{k=1}^{m} p_{sk}^* \mathbf{z}_n^* (\lambda_k^{(n)}) \quad (s = 1, \cdots, m; 0 < m < n), \tag{4.15}$$

it is the trend SIML-smoothing value for $\mathbf{r}_s^{*(m,n)}$. It is $\hat{\mathbf{X}}_n^*(m)$ $(= \mathbf{C}_n^{\prime}\mathbf{R}_n^*(m))$. and

$$\mathbf{r}_{s-1}^{*(m,n)} = \sum_{k=1}^{m} p_{sk}^* \mathbf{z}_n^{**}(\lambda_k^{(n)}) \ (s = 1, \cdots, m; 0 < m < n), \tag{4.16}$$

where $\mathbf{z}^{**(n)}(\lambda_k^{(n)})$ are constructed from the $n \times p$ hidden states matrix $\mathbf{X}_n^*$, rather than the observed $n \times p$ matrix data $\mathbf{Y}_n^*$. Hence, it is the same as the element of $\mathbf{C}_n^{\prime -1}\hat{\mathbf{X}}_n^*(m)$, and for $\lambda_m^{(n)} = m/n$ in the frequency domain, it is a discrete version of

$$\hat{\mathbf{r}}_{s-1}^{*(n)}(\lambda_m^{(n)}) = \int_0^{\lambda_m^{(n)}} \sin[2\pi\lambda\, s] dB_n(\lambda) \ . \tag{4.17}$$

Then, this term corresponds to the element of $\mathbf{C}_n^{\prime -1}\hat{\mathbf{X}}_n^*(m)$, and has the corresponding (continuous) version in the frequency domain.
Similarly, $\hat{\mathbf{r}}_{s-1}^{*m_1,m_2,n)} = \sum_{k=m_1+1}^{m_1+m_2} p_{sk} \Delta_\lambda \mathbf{z}^{*(n)}(\lambda_k^{(n)}) = \hat{\mathbf{r}}_{s-1}^{*(m_2,n)} - \hat{\mathbf{r}}_{s-1}^{*(m_1,n)}$ $(s = 1, \cdots, m; 0 < m_1 < m_2 < n)$ can be regarded as a discrete version of

$$\hat{\mathbf{r}}_{s-1}^{*(n)}(\lambda_{m_1}^{(n)}, \lambda_{m_2}^{(n)}) = \int_{\lambda_{m_1}^{(n)}}^{\lambda_{m_2}^{(n)}} \sin[2\pi\lambda\, s] dB_n(\lambda) \ . \tag{4.18}$$

5. Numerical Example

We illustrate the use of the SIML-forward smoothing and SIML-backward smoothing for real data. We have used the monthly US Manufacturers' New Orders Data within 2010-2020 because it is known that this time series data has trend, wild seasonal fluctuation and noise components.

The red curve in Figure 1 shows the forward smoothing, given the first observation as the initial condition with $m = 5$. The green curve in the figure shows T_1^* as the limit of the forward-backward iterations. The violet curve in the figure shows the two-step forward filtering with $m_1 = 15$ (first smoothing) and $m = 5$ (second smoothing). The blue curve in Figure 2 shows the backward smoothing given the last observation as the initial condition with $m = 5$. The sky-blue curve in the figure show T_2^* as the limit of the backward-backward iterations. The violet curve in the figure shows the two-step backward filtering with $m_1 = 15$ (first smoothing) and $m = 5$ (second smoothing).

As we have expected, the initial values of the both forward and backward smoothers have significant effects around the initial values at which we start smoothing. The effects of choosing the initial value become negligible either repeating smoothing (or filtering) and multi-step smoothing. In particular, the resulting differences in two procedures after a few steps are small for practical purposes.

6. Conclusions

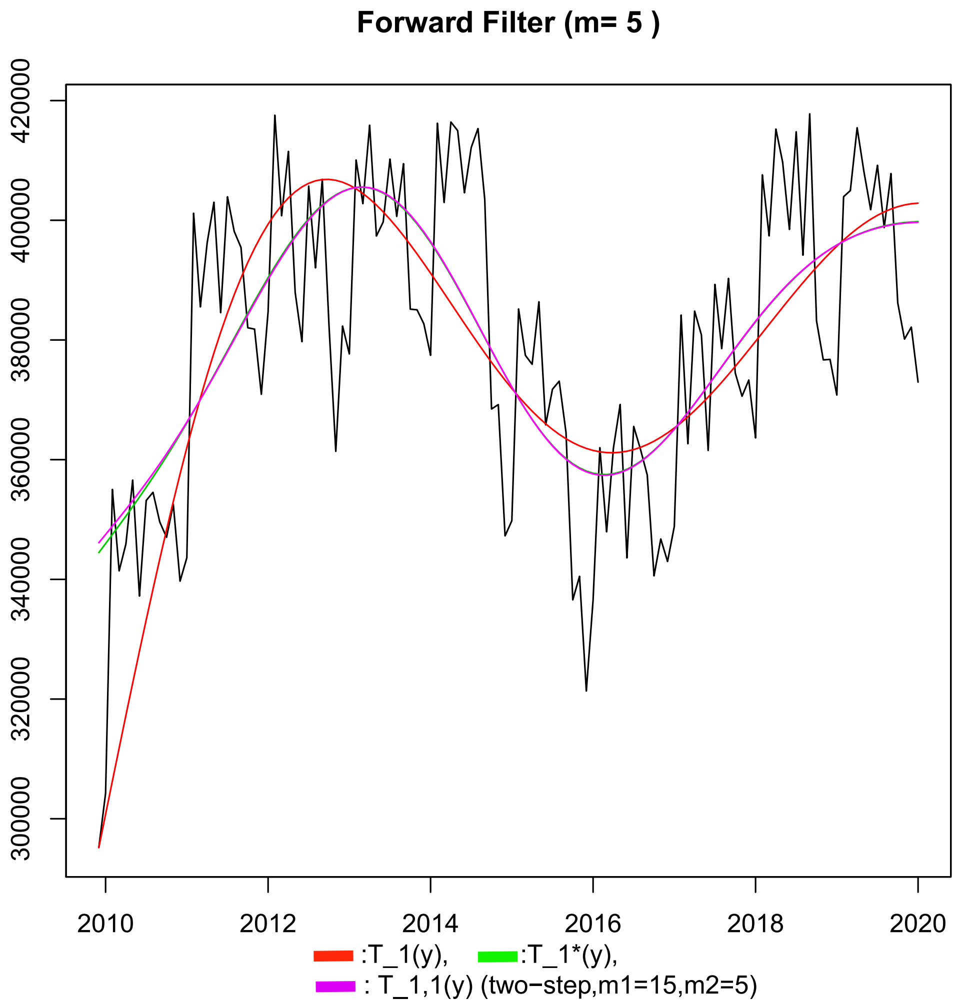

Figure 1: Forward filtering results for monthly US Manufacturers' New Orders from 2010 to 2020. (https://www.census.gov/manufacturing/m3/index.html)

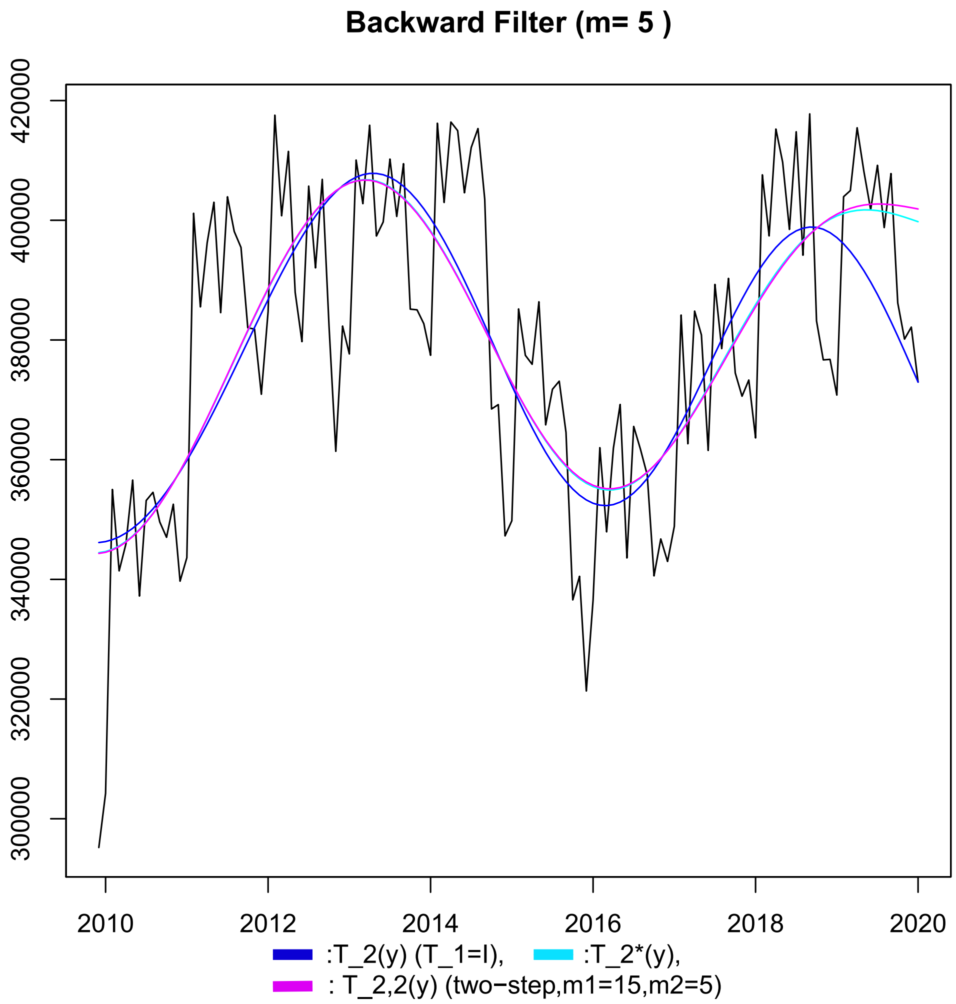

Figure 2: Backward filtering results for monthly US Manufacturers' New Orders from 2010 to 2020. (https://www.census.gov/manufacturing/m3/index.html)

When the observed non-stationary multivariate time series contain noises, undoing the effects of trends and noises may be difficult. This study follows Kunitomo and Sato (2021), who investigated a new statistical smoothing procedure to decompose time series into non-stationary trend, seasonal, and stationary noise (or measurement errors) components. The resulting smoothing or filtering method for the non-stationary multivariate series is simple and free from the underlying distributions of the noise and state vector. Therefore, it is robust against possible misspecification in the non-stationary multivariate time series.

Several interesting problems were developed by the approach proposed in this study. Our method framework presents some earlier studies on the filtering methods in the time and frequency domains. Although our method is a non-parametric smoothing method, a close relationship is observed with the existing smoothing and filtering methods such as Decomp by Kitagawa (2010), which followed Akaike (1980) (refer to in Kunitomo and Sato (2021)). This problem is currently under investigation. Furthermore, there are likely to have many empirical applications, which will be examined in future studies, including the one in Sato and Kunitomo Sato (2021).

References

[1] Akaike, H. (1980), "Seasonal Adjustment by a Bayesian Modeling," *Journal of Time Series Analysis*, 1, 1-13.

[2] Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, Wiley.

[3] Brillinger, D. and M. Hatanaka (1969) "An Harmonic Analysis of Nonstationary Multivariate Economic Processes," *Econometrica*, Vol. 35, 131-141.

[4] Brillinger, D. (1980) *Time Series : Data Analysis and Theory*, Expanded Edition, Holden-Day.

[5] Brockwell, P. and R. Davis (1990), *Time Series : Theory and Methods*, 2nd Edition, Wiley.

[6] Doob, J.L (1953), *Stochastic Processes*, Wiley.

[7] Harvey, A. and T. Trimbur (2008), "Trend Estimation and The Hodorick-Prescott Filter," Journal of Japan Statistical Society, 38-1, 41-49.

[8] Kitagawa, G. (2010), *Introduction to Time Series Analysis*, CRC Press.

[9] Kunitomo, N. , S. Sato and D. Kurisu (2018), *Separating Information Maximum Likelihood Estimation for High Frequency Financial Data*, Springer.

[10] Kunitomo, N. and S. Sato (2021), "A Robust-filtering Method for Noisy Non-Stationary Multivariate Time Series with Econometric Applications," Japanese Journal of Statistics and Data Science, 4, 373-410. https://doi.org/10.1007/s42081-020-00102-y.

[11] Kunitomo, N., N. Awaya and D. Kurisu (2020), "Some Properties of Estimation Methods for Structural Relationships in Non-stationary Errors-in-Variables Models," Japanese Journal of Statistics and Data Science, 3, 73-101.

[12] Nishimura, G. K., S. Sato, A. Takahashi (2019), "Term Structure Models During the Global Financial Crisis: A Parsimonious Text Mining Approach," *Asic-Pacific Financial Markets*.

[13] Sato, S. and Kunitomo, N. (2021), "Frequency Regression and Smoothing for Noisy Non-stationary Time Series," MIMS-RBP Statistics & Data Science Series (SDS-20), http://www.mims.meiji.ac.jp/publications/2021-ds/SDS-20.pdf.

APPENDIX : Mathematical Derivations

We present here some details of derivations that we have omitted in the previous sections. Most of our derivations is to apply trigonometric relations, which are mathematically elementary and straightforward. Hence, we show only the essential parts of derivations and prepare a lemma on the characteristic roots and eigen vectors of a patterned matrix, Then, we show the proof of theorems.

Lemma A.1 : (i) Define an $n \times n$ matrix $\mathbf{A}_n^*$ by

$$\text{(A.1)} \qquad \mathbf{A}_n^* = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ 0 & 1 & 0 & 1 & \cdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & \cdots & 0 & 1 & 1 \end{pmatrix} .$$

Then, $\cos \pi(\frac{2k-1}{2n+1})$ $(k = 1, \cdots, n)$ are eigen-values of $\mathbf{A}_n^*$, and the eigen-vectors are

$$\text{(A.2)} \qquad \begin{bmatrix} \sin[\pi(\frac{2k-1}{2n+1})1] \\ \sin[\pi(\frac{2k-1}{2n+1})2] \\ \vdots \\ \sin[\pi(\frac{2k-1}{2n+1})n] \end{bmatrix} \quad (k = 1, \cdots, n).$$

(ii) We have the spectral decomposition

$$\text{(A.3)} \qquad \mathbf{C}_n'^{-1}\mathbf{C}_n^{-1} = \mathbf{P}_n^{*'}\mathbf{D}_n\mathbf{P}_n^* = 2\mathbf{I}_n - 2\mathbf{A}_n^* ,$$

where $\mathbf{P}^{*'}$ is the matrix consisting of eigen-vectors in (A.2), $\mathbf{D}_n$ is a diagonal matrix with the $k-$th element

$$\text{(A.4)} \qquad d_k = 2\left[1 - \cos(\pi(\frac{2k-1}{2n+1}))\right] \quad (k = 1, \cdots, n) ,$$

$$\text{(A.5)} \qquad \mathbf{C}_n'^{-1} = \begin{pmatrix} 1 & -1 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and the $(k, j)-$the element of $\mathbf{P}_n^* = (p_{kj}^*)$ is given by

$$\text{(A.6)} \qquad p_{kj}^* = \sqrt{\frac{2}{n + \frac{1}{2}}} \sin\left[\frac{2\pi}{2n+1}(k - \frac{1}{2})\, j\right] .$$

Proof of Lemma A.1 : (i) Let $\mathbf{A}_n^* = (a_{ij}^*)$ $(i,j = 1,\cdots,n)$ and an $n \times 1$ vector $\mathbf{x} = (x_t)$ $(t = 1,\cdots,n)$ satisfying $\mathbf{A}_n^*\mathbf{x} = \lambda\mathbf{x}$. Then,

$$\text{(A.7)} \qquad \frac{x_2}{2} = \lambda x_1 \ ,$$

$$\text{(A.8)} \qquad \frac{x_{t-1}+x_{t+1}}{2} = \lambda x_t \ (t = 2,\cdots,n-1) \ ,$$

$$\text{(A.9)} \qquad \frac{1}{2}[x_{n-1}+x_n] = \lambda x_n \ .$$

Let ξ_i $(i = 1,2)$ be the solutions of $\xi^2 - 2\lambda\xi + 1 = 0$. Because $2\lambda = \xi_1 + \xi_2$ and $\xi_1\xi_2 = 1$, we have the solution as $x_t = c_1\xi_1^t + c_2\xi_1^{-t}$ $(t = 1,\cdots,n)$ and c_i $(i = 1)$ are real constants. The first equation implies $0 = c_1\xi_1^2 + c_2\xi_1^{-2} - (\xi_1+\xi_1^{-1})(c_1\xi_1 + c_2\xi_1^{-1})$, and $c_1 + c_2 = 0$. Then, we find that $x_t = c_1[\xi_1^t - \xi_1^{-t}]$, and the third equation implies $(\xi_1^{2n+1}+1)(1-\xi_1) = 0$. Because $\xi_1 \neq 1$, we find that $\xi_1^{2n+1} = -1 = e^{\pi i(2k-1)}$ for any positive integer k.
Then,

$$\text{(A.10)} \qquad \lambda_k = \cos[\pi\frac{2k-1}{2n+1}] \ (k = 1,\cdots,n) \ .$$

By taking $c_1 = (1/2i)$, the elements of the characteristic vectors of $\mathbf{A}_n^*$ with $\cos[\pi(2k-1)/(2n+1)]$ are

$$\text{(A.11)} \qquad x_t = \frac{1}{2i}\left[\xi_1^t - \xi_1^{-t}\right] = \sin\left[\pi\frac{2k-1}{2n+1}t\right] \ .$$

(ii) The rest of the proof involves the standard arguments of spectral decomposition in linear algebra. **Q.E.D.**

Proof of Theorem 3.1 :

(i) We consider the case of T_{2k+1} $(k \geq 1)$. By using the recursive relations, for $k \geq 1$ we can represent

$$\text{(A.12)} \qquad T_{2k+1} = A_1^{(m,n)} + A_2^{(n)}T_{2(k-1)+1} \ ,$$

where an $n \times n$ matrix $A_2^{(m,n)}$ is defined by

$$\text{(A.13)} \ A_2^{(m,n)} = (\mathbf{I}_n - \mathbf{C}_n\mathbf{P}_n\mathbf{J}_m^{'}\mathbf{J}_m\mathbf{P}_n\mathbf{C}^{-1})\mathbf{1}_n\mathbf{e}_1^{'}(\mathbf{I}_n - \mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{J}_m^{'}\mathbf{J}_m\mathbf{P}_n^{*}\mathbf{C}^{'-1})\mathbf{1}_n\mathbf{e}_n^{'} \ .$$

Then, we consider the characteristic roots of the coefficient matrix $\mathbf{A}_2^{(m,n)}$. Because the rank of $\mathbf{A}_2^{(n)}$ is one, there are $n-1$ zero roots and one non-zero root, which is

$$\begin{aligned}\text{(A.14)}\, a_{2n} &= \mathbf{e}_n^{'}(\mathbf{I}_n - \mathbf{C}_n\mathbf{P}_n\mathbf{J}_m^{'}\mathbf{J}_m\mathbf{P}_n\mathbf{C}^{-1})\mathbf{1}_n\mathbf{e}_1^{'}(\mathbf{I}_n - \mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{J}_m^{'}\mathbf{J}_m\mathbf{P}_n^{*}\mathbf{C}^{'-1})\mathbf{1}_n \\ &= [1 - \mathbf{1}_n^{`}\mathbf{P}_n\mathbf{J}_m^{'}\mathbf{J}_m\mathbf{P}_n\mathbf{e}_1][1 - \mathbf{1}_n^{`}\mathbf{P}_n^{*'}\mathbf{J}_m^{'}\mathbf{J}_m\mathbf{P}_n^{*}\mathbf{e}_n] \ .\end{aligned}$$

(We have used that $\mathbf{e}_n^{'}\mathbf{C}_n = \mathbf{1}_n^{`}$ and $\mathbf{e}_1^{'}\mathbf{C}_n^{'} = \mathbf{1}_n^{`}$.)
By using the relation $1 - \mathbf{1}_n^{'}\mathbf{P}_n\mathbf{J}_m^{'}\mathbf{J}_m\mathbf{P}_n\mathbf{e}_1 = \mathbf{1}_n^{'}\mathbf{P}_n\mathbf{J}_{n-m}^{'}\mathbf{J}_{n-m}\mathbf{P}_n\mathbf{e}_1$ for $\mathbf{J}_{n-m} =$

$(\mathbf{O}, \mathbf{I}_{n-m})$ ($(n-m) \times [m+(n-m)]$ matrix), we evaluate two terms in (A.14). The first term of (A.14) becomes

$$[\sqrt{\frac{2}{n+\frac{1}{2}}}]^2 \sum_{k=m+1}^{n} [\sum_{j=1}^{n} \cos \frac{2\pi}{2n+1}(j-\frac{1}{2})(k-\frac{1}{2})] \times \cos \frac{2\pi}{2n+1}(k-\frac{1}{2})(1-\frac{1}{2})] ,$$

which is less than 1. It is because, by using Lemma 5.1 of Kunitomo et al. (2018), the above term is

$$[\frac{2}{2n+1}] \sum_{k=m+1}^{n} [[\frac{\sin \frac{2\pi}{2n+1}(k-\frac{1}{2})n}{\sin \frac{\pi}{2n+1}(k-\frac{1}{2})}] \times \cos \frac{2\pi}{2n+1}(k-\frac{1}{2})(1-\frac{1}{2})]$$

and the relation

$$\sin \frac{\pi}{2n+1}(k-\frac{1}{2})[2n+1-1] = \sin \pi(k-\frac{1}{2}) \cos \frac{\pi}{2n+1}(k-\frac{1}{2}) ,$$

it becomes

$$[\frac{2}{2n+1}] \sum_{k=m+1}^{n} \sin \pi(k-\frac{1}{2}) \times [\frac{[\cos \frac{\pi}{2n+1}(k-\frac{1}{2})]^2}{\sin \frac{\pi}{2n+1}(k-\frac{1}{2})}] .$$

Because $\sin \pi(k-\frac{1}{2})$ takes $+1$ and -1 alternatively, we evaluate the difference of

$$\frac{[\cos \frac{\pi}{2n+1}(k-\frac{1}{2})]^2}{\sin \frac{\pi}{2n+1}(k-\frac{1}{2})} - \frac{[\cos \frac{\pi}{2n+1}(k-1-\frac{1}{2})]^2}{\sin \frac{\pi}{2n+1}(k-1-\frac{1}{2})} \sim \frac{[\cos \frac{\pi}{2n+1}(k-\frac{1}{2})]^2[1-\cos \frac{\pi}{2n+1}]}{\sin \frac{\pi}{2n+1}(k-\frac{1}{2})} .$$

We can take $n > n_0$ such that $\sin \frac{\pi}{2n+1}$ and $1-\cos \frac{\pi}{2n+1}$ being sufficient small. Then, each term becomes small, and then $\mathbf{1}_n^{'}\mathbf{P}_n\mathbf{J}_{n-m}^{'}\mathbf{J}_{n-m}\mathbf{P}_n\mathbf{e}_1$ is less than one. Similarly, for the second term of (A.14), we use

$$[1-\mathbf{1}_n^{'}\mathbf{P}_n^{*'}\mathbf{J}_m^{'}\mathbf{J}_m\mathbf{P}_n^{*}\mathbf{e}_n] = \mathbf{1}_n^{'}\mathbf{P}_n^{*'}\mathbf{J}_{n-m}^{'}\mathbf{J}_{n-m}\mathbf{P}_n^{*}\mathbf{e}_n .$$

Then, the second term of (A.14) becomes

$$[\sqrt{\frac{2}{n+\frac{1}{2}}}]^2 \sum_{k=1}^{n} \sum_{j=m+1}^{n} [\sin \frac{2\pi}{2n+1}(k-\frac{1}{2})j \sin \frac{2\pi}{2n+1}(n-\frac{1}{2})j ,$$

which is less than 1. In this evaluation, we have utilized the relation that

(A.15)
$$\sum_{k=1}^{n} \sin \frac{2\pi}{2n+1}(k-\frac{1}{2})j = \frac{1}{2i}\frac{e^{i\frac{2\pi}{2n+1}jn}+e^{-i\frac{2\pi}{2n+1}jn}-2}{e^{i\frac{2\pi}{2n+1}j\frac{1}{2}}-e^{-i\frac{2\pi}{2n+1}j\frac{1}{2}}} = \frac{1}{2}\frac{1-\cos\frac{2\pi}{2n+1}jn}{\sin\frac{2\pi}{2n+1}j\frac{1}{2}}$$

and the elementary relation on trigonometric functions that for $2(n-1/2)=(2n+1)-2$

$$\sin\frac{2\pi}{2n+1}(n-\frac{1}{2})j=[-\cos\pi j]\sin\frac{2\pi}{2n+1}j=2[-\cos\pi j]\sin\frac{\pi}{2n+1}j\cos\frac{\pi}{2n+1}j\ .$$

Because each term of (A.14) is less than one, we have $|a_{2n}|<1$. Then, by using (A.12), we have convergence of T_{2k+1} as $k\to\infty$.
(ii) We can apply the similar arguments to T_{2k} $(k\geq 1)$. By using the recursive relations, for $k\geq 1$ we can represent

$$T_{2k}=A_{1*}^{(m,n)}+A_{2*}^{(n)}T_{2(k-1)+1}\ , \tag{A.16}$$

where $A_{1*}^{(m,n)}$ and $A_{2*}^{(m,n)}$ are $n\times n$ matrices as defined in Theorem 3.1. By evaluating the eigenvalues of $A_{2*}^{(m,n)}$, we find that the absolute value of engenvalues are less than one, and we have convergence of T_{2k} as $k\to\infty$.
(**Q.E.D.**)

Proof of Theorem 3.2 :

(i) From (3.8) and (2.9), we set

$$h(a,b,n)=[\frac{2n+1}{4}]h_{ab}^{(n)}=\sum_{j=1}^{m}\cos\frac{2\pi}{2n+1}(a-\frac{1}{2})(j-\frac{1}{2})\cos\frac{2\pi}{2n+1}(b-\frac{1}{2})(j-\frac{1}{2}) \tag{A.17}$$

and

$$f(a,b,n)=[\frac{2n+1}{4}]f_{ab}^{(n)}=\sum_{j=1}^{m}\cos\frac{2\pi}{2n+1}(a-\frac{1}{2})j\cos\frac{2\pi}{2n+1}(b-\frac{1}{2})j\ . \tag{A.18}$$

Then, we evaluate $h(a,b,n)-f(a,b,n)$ by using elementary trigonometric relations such as $2\cos\theta_1\cos\theta_2=\cos(\theta_1+\theta_2)+\cos(\theta_1-\theta_2)$, $\cos\theta_1 j=\cos\theta_1(j-\frac{1}{2})\cos(\theta_1/2)-\sin\theta_1(j-\frac{1}{2})\sin(\theta_1/2)$ for some θ_1 and θ_2. Also we use the relation $\sum_{k=1}^{m}\cos[2\pi/(2n+1)]l(k-1/2)=(1/2)[\sin(2\pi ml/(2n+1))]/[\sin(2\pi l/(2n+1))]$ for positive integers l,m (Lemma 5.1 of Kunitomo, Sato and Kurisu (2018)). There are four terms to be evaluated. One (typical) term of the resulting many terms is $(1/2)[1-\cos 2\pi/(2n+1)(a+b-1)/2]\sum_{j=1}^{m}\cos 2\pi/(2n+1)(a+b-1)(j-1/2)$. If $(a+b-1)/n\to 0$, then $[1-\cos 2\pi/(2n+1)(a+b-1)/2]\to 0$. If $(a+b-1)/n\to c$ (a non-zero constant), $\sum_{j=1}^{m}\cos 2\pi/(2n+1)(a+b-1)(j-1/2)$ converges to a constant. In both cases, the term $(1/m)[1-\cos 2\pi/(2n+1)(a+b-1)/2]\sum_{j=1}^{m}\cos 2\pi/(2n+1)(a+b-1)(j-1/2)$ becomes small arbitrary as $m\to\infty$.
After these (straightforward) calculations, we find that $(1/m)[h(a,b,n)-f(a,b,n)]\to 0$ as $n\to\infty$ and $n/m\to\infty$ $(m=m_n)$.
(ii) We take the trace operation to find that

$$\mathrm{Tr}[\mathbf{H}_n^{(0)}-\mathbf{F}_n^{*}]=\mathrm{Tr}[\mathbf{H}_n-\mathbf{F}_n]\ . \tag{A.19}$$

From (A.17) and (A.18), by using the simple relation $2\cos\theta_1\cos\theta_2 = \cos(\theta_1+\theta_2) + \cos(\theta_1-\theta_2)$, we have

$$\text{(A.20)} \qquad 2h(a,a,n) = \sum_{j=1}^{m}\left[\cos\frac{2\pi}{2n+1}(2a-1)(j-\frac{1}{2})+1\right] ,$$

$$\text{(A.21)} \qquad 2f(a,a,n) = \sum_{j=1}^{m}\left[\cos\frac{2\pi}{2n+1}(2a-1)j+1\right] .$$

We use Lemma 5.1 of Kunitomo, Sato and Kurisu (2018) when $m=n$ to find

$$\text{(A.22)} \qquad \sum_{k=1}^{n}\cos[\frac{2\pi}{2n+1}]l(k-\frac{1}{2}) = (-\frac{1}{2})\cos\pi l$$

for any positive intege l. Then

$$\begin{aligned} 2\sum_{a=1}^{n}[h(a,a,n)-f(a,a,n)] &= \sum_{j=1}^{m}\sum_{a=1}^{n}[\cos\frac{2\pi}{2n+1}(a-\frac{1}{2})(2j-1)-\cos\frac{2\pi}{2n+1}(a-\frac{1}{2})2j] \\ &= \sum_{j=1}^{m}(-\frac{1}{2})[\cos\pi(2j-1)-\cos\pi(2j)] = m . \end{aligned}$$

Because

$$\text{(A.23)} \qquad \mathrm{Tr}[\mathbf{H}_n-\mathbf{F}_n] = \frac{4}{2n+1}\sum_{a=1}^{n}[h(a,a,n)-f(a,a,n)] ,$$

we have the result.
(**Q.E.D.**)

Proof of Theorem 3.3 :

The proof is basically the same as Theorem 3.1. We replace $\mathbf{Q}_n^{(m)} = \mathbf{J}_m^{'}\mathbf{J}_m$ by $\mathbf{Q}_n^{(m_1,m_2)} = \mathbf{J}_{m_1,m_2}^{'}\mathbf{J}_{m_1,m_2}$.
We illustrate an example and consider $T_{2k+1} = A_1^{(m_1,m_2,,n)} + A_2^{(m_1,m_2,n)}T_{2k-1}$. Then, the non-zero eigenvalue of $A_2^{(m_1,m_2,n)}$ is

$$\begin{aligned} a_{2n} &= \mathbf{e}_n^{'}(\mathbf{I}_n-\mathbf{C}_n\mathbf{P}_n\mathbf{J}_{m_1,m_2}^{'}\mathbf{J}_{m_1,m_2}\mathbf{P}_n\mathbf{C}^{-1})\mathbf{1}_n\mathbf{e}_1^{'}(\mathbf{I}_n-\mathbf{C}_n^{'}\mathbf{P}_n^{*'}\mathbf{J}_{m_1,m_2}^{'}\mathbf{J}_{m_1,m_2}\mathbf{P}_n^{*}\mathbf{C}^{'-1})\mathbf{1}_n \\ &= [1-\mathbf{1}_n^{'}\mathbf{P}_n\mathbf{J}_{m_1,m_2}^{'}\mathbf{J}_{m_1,m_2}\mathbf{P}_n\mathbf{e}_1][1-\mathbf{1}_n^{'}\mathbf{P}_n^{*}\mathbf{J}_{m_1,m_2}^{'}\mathbf{J}_{m_1,m_2}\mathbf{P}_n^{*}\mathbf{e}_n] . \end{aligned}$$

In this case, we use the relation

$$\begin{aligned} [1-\mathbf{1}_n^{'}\mathbf{P}_n\mathbf{J}_{m_1,m_2}^{'}\mathbf{J}_{m_1,m_2}\mathbf{P}_n\mathbf{e}_1] &= \mathbf{1}_n^{'}\mathbf{P}_n\mathbf{J}_{m_1}^{'}\mathbf{J}_{m_1}\mathbf{P}_n\mathbf{e}_1 \\ &+ \mathbf{1}_n^{'}\mathbf{P}_n\mathbf{J}_{n-m_1-m_2,n-m_1-m_2}^{'}\mathbf{J}_{n-m_1-m_2,n-m_1-m_2}\mathbf{P}_n\mathbf{e}_1 . \end{aligned}$$

Then, by using the similar arguments as the proof of Theorem 3.1, we find that the absolute values of the eigenvalues of $\mathbf{A}_2^{(m_1,m_2,n}$ is less than one. We use a similar argument to show that non-zero eigenvalue of $\mathbf{A}_{2*}^{(m_1,m_2,n}$ is less than one, and then, we have the convergence of the repeated smoothing procedures.
(**Q.E.D.**)