SDS-3

Some Properties of Estimation Methods for Structural Relationships in Non-stationary Errors-in-Variables Models

Naoto Kunitomo, Naoki Awaya
and Daisuke Kurisu

November 2017

Statistics & Data Science Series back numbers:
http://www.mims.meiji.ac.jp/publications/datascience.html

Some Properties of Estimation Methods for Structural Relationships in Non-stationary Errors-in-Variables Models *

Naoto Kunitomo †
Naoki Awaya ‡
and
Daisuke Kurisu §

September 20, 2017

Abstract

Kunitomo and Sato (2017) developed a new statistical method called the Separating Information Maximum Likelihood (SIML) estimation for the multivariate non-stationary errors variables models. In this paper we compare the SIML estimation and the maximum likelihood (ML) estimation when there are non-stationary trends and noise components in the underlying multivariate time series. We have found that the Gaussian likelihood function can have some peculiar shape and the computation of ML estimation can be unstable in some cases. The SIML estimation has the asymptotic robust properties under general conditions of the existence of fourth order moments and it does not have any computational problem. We show the asymptotic properties of the ML estimation and SIML estimation as well as their finite sample properties.

Key Words

Non-stationary economic time series, Errors-variables models, Co-integrated trends, $\mathbf{K}_n-$transfomation, Likelihood Function, Maximum Likelihood (ML), Separating Information Maximum Likelihood (SIML), Asymptotic Robustness

---

*This is a preliminary memorandum and the research has been supported by JSPS Grant-in-Aid for Scientific Research JP17H02513.

†School of Potitical Science and Economics, Meiji University, Kanda-Surugadai 1-1, Chiyoda-ku 101-8301, Tokyo, JAPAN, naoto.kunitomo@gmail.com

‡Graduate School of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1, Tokyo 113-0033

§Graduate School of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1, Tokyo 113-0033, JAPAN

1. Introduction

There is a vast amount of published research on the use of statistical time series analysis of macroeconomic time series. One important distinction of macroeconomic time series from the standard time series analysis in other areas has been the mixture of non-stationarity and measurement errors including apparent seasonality. The measurement errors in economic times series are often essential in the published official macro-economic data, but enough attention has not been in the analysis of non-stationary multivariate time series analysis.

In this regard, Kunitomo and Sato (2017) have developed a new statistical method called the Separating Information Maximum Likelihood (SIML) estimation for the multivariate non-stationary errors-in-variables models. Earlier and related literature on the non-stationary economic time series are Engle and Granger (1987) and Johansen (1995), which have dealt with the econometric modeling of multivariate non-stationary and stationary time series and developed the notion of co-integration, but they did not have paid much attention on *measurement errors*. The problem discussed by Kunitomo and Sato (2017) is related to their work, but it has different aspects due to the fact that its main focus was on the non-stationarity, seasonality and measurement error in the non-stationary errors-in-variables models. Many economists tend to use the official macro-economic data, but then they should have taken more attention on the measurement errors and official seasonal adjustments. There are important consequences by using economic data with the measurement errors, which are common in many published macro-economic time series by central governments.

On the other hand, in the literature of statistical time series analysis, the state space modeling for non-stationary time series has been developed by Akaike (1989) and Kitagawa (2010) and there have been many applications in many fields including control engineering and statistical seismology already reported (see Ohtsu, Peng and Kitagawa (2015) for instance). Their statistical method may look different from time series econometrics at the first glance. By looking their method carefully and comparing with the econometric analysis of non-stationary multivariate time series, the underlying statistical problems are quite similar with respect to the non-stationarity and measurement errors. Then the study of Kunitomo and Sato (2017) can be regarded as an investigation of statistical inference problem of state space modeling, which is related to an earlier work by Chiang, Jiang and Park (2008).

The main purpose of this paper is to compare the the SIML estimation and the maximum likelihood (ML) estimation, which are two different methods to estimate the non-stationary errors-in-variables models when there are non-stationary trends and noise components. In this paper we will report that the Gaussian likelihood function show some peculiar shape and the computation of the ML estimation can be unstable in some cases. Also when there are co-integrated relations among trends with the rank being smaller than the dimension of observations, there could be an additional computational problem. On the other hand, the separating information maximum likelihood (SIML) method gives an alternative way to overcome the underlying difficulty of computation while it has reasonable statistical properties. The SIML estimation has the asymptotic robust properties under general conditions of existence of second order moments. We show the asymptotic properties as well as its finite sample properties of the ML estimation and SIML estimation as well as simulations.

In Section 2 we will present a general formulation of the problem and give simple examples to illustrate the problem in this study. Then In Section 3, we will discuss the simple one-dimensional case and the non-stationary model has the random walk plus noise and we will develop one common (non-stationary) factor case in multivariate errors-in-variables models. Section 4 we investigate the Gaussian likelihood function and its shape. We give the consistency result of the ML estimation. It seems to be new although it is not surprising result. In Section 5, we will discuss some extensions and then present some concluding remarks in Section 6. Some mathematical derivations will be given in Appendix.

2. The Model

Let y_{ij} be the $i-$th observation of the $j-$th time series at i for $i = 1, \cdots, n; j = 1, \cdots, p$. We set $\mathbf{y}_i = (y_{1i}, \cdots, y_{pi})^{'}$ be a $p \times 1$ vector and $\mathbf{Y}_n = (\mathbf{y}_i^{'})\ (= (y_{ij}))$ be an $n \times p$ matrix of observations and we denote $\mathbf{y}_0$ as the initial $p \times 1$ vector. We consider the situation when the underlying non-stationary trends $\mathbf{x}_i\ (= (x_{ji}))\ (i = 1, \cdots, n)$ are not necessarily the same as the observed time series and let $\mathbf{v}_i^{'} = (v_{1i}, \cdots, v_{pi})$ be the vectors of the noise components, which are independent of $\mathbf{x}_i$. Then we use the additive decomposition form

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

where the trend components $\mathbf{x}_i\ (i = 1, \cdots, n)$ satisfies

$$\Delta \mathbf{x}_i = (1 - \mathcal{L})\mathbf{x}_i = \mathbf{w}_i^{(x)} \tag{2.2}$$

with $\mathcal{L}\mathbf{x}_i = \mathbf{x}_{i-1}$, $\Delta = 1 - \mathcal{L}$, $\mathcal{E}(\mathbf{w}_i^{(x)}) = \mathbf{0}$, $\mathcal{E}(\mathbf{w}_i^{(x)}\mathbf{w}_i^{(x)'}) = \mathbf{\Sigma}_x$, and a sequence of stationary components satisfies $\mathbf{v}_i\ (i = 1, \cdots, n)$ with $\mathcal{E}(\mathbf{v}_i\mathbf{v}_i^{'}) = \mathbf{\Sigma}_v$ and

$$\mathbf{v}_i = \sum_{j=-\infty}^{\infty} \mathbf{C}_j \mathbf{e}_{i-j}\ , \tag{2.3}$$

with absolutely summable coefficients $\mathbf{C}_j$ and a sequence of i.i.d. random vectors with $\mathcal{E}(\mathbf{e}_i) = \mathbf{0}$, $\mathcal{E}(\mathbf{e}_i\mathbf{e}_i^{'}) = \mathbf{\Sigma}_e$.

We assume that $\mathbf{w}_i^{(x)}$ and $\mathbf{e}_i$ are the sequence of i.i.d. random vectors with $\mathbf{\Sigma}_e$ being positive-semi-definite, and the random vectors $\mathbf{w}_i^{(x)}$ and $\mathbf{e}_i$ are mutually independent. When $\mathbf{v}_i = \mathbf{e}_i$, we can interpret that it is a sequence of independent measurement errors. The present additive decomposition is similar to the one given by Kitagawa and Gersch (1984) and Kitagawa (2010).

In order to develop the SIML estimation, we first consider the situation when $\Delta \mathbf{x}_i$ and $\mathbf{v}_i$ $(i = 1, \cdots, n)$ are mutually independent and each of the component vectors are independently, identically, and normally distributed as $N_p(\mathbf{0}, \mathbf{\Sigma}_x)$ and $N_p(\mathbf{0}, \mathbf{\Sigma}_v)$, respectively. We use an $n \times p$ matrix $\mathbf{Y}_n = (\mathbf{y}_i^{'})$ and consider the distribution of $np \times 1$ random vector $(\mathbf{y}_1^{'}, \cdots, \mathbf{y}_n^{'})^{'}$. Given the initial condition $\mathbf{y}_0$, we have

$$\text{vec}(\mathbf{Y}_n) \sim N_{n \times p}\left(\mathbf{1}_n \cdot \mathbf{y}_0^{'}, \mathbf{I}_n \otimes \mathbf{\Sigma}_v + \mathbf{C}_n \mathbf{C}_n^{'} \otimes \mathbf{\Sigma}_x\right) , \tag{2.4}$$

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

$$\mathbf{C}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 0 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}_{n \times n} . \tag{2.5}$$

Then, given the initial condition $\mathbf{y}_0$ the conditional maximum likelihood (ML) estimator can be defined as the solution of maximizing the conditional log-likelihood function except a constant as

$$\begin{aligned} L_n^* &= \log |\mathbf{I}_n \otimes \mathbf{\Sigma}_v + \mathbf{C}_n \mathbf{C}_n^{'} \otimes \mathbf{\Sigma}_x|^{-1/2} \\ &\quad -\frac{1}{2}[vec(\mathbf{Y}_n - \bar{\mathbf{Y}}_0)^{'}]^{'}[\mathbf{I}_n \otimes \mathbf{\Sigma}_v + \mathbf{C}_n \mathbf{C}_n^{'} \otimes \mathbf{\Sigma}_x]^{-1}[vec(\mathbf{Y}_n - \bar{\mathbf{Y}}_0)^{'}] \quad , \end{aligned}$$

where

$$\bar{\mathbf{Y}}_0 = \mathbf{1}_n \cdot \mathbf{y}_0^{'} . \tag{2.6}$$

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

$$\mathbf{Z}_n = \mathbf{K}_n \left(\mathbf{Y}_n - \bar{\mathbf{Y}}_0\right) , \mathbf{K}_n = \mathbf{P}_n \mathbf{C}_n^{-1} , \tag{2.7}$$

where

$$\mathbf{C}_n^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}_{n \times n} , \tag{2.8}$$

and

$$\mathbf{P}_n = (p_{jk}^{(n)}) \ , \ p_{jk}^{(n)} = \sqrt{\frac{2}{n+\frac{1}{2}}} \cos\left[\frac{2\pi}{2n+1}(k-\frac{1}{2})(j-\frac{1}{2})\right] \ . \tag{2.9}$$

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 the conditional likelihood function given the initial condition is proportional to

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

where

$$a_{kn}^{*} \ (= d_k) \ = 4\sin^2\left[\frac{\pi}{2}\left(\frac{2k-1}{2n+1}\right)\right] \ (k=1,\cdots,n) \ . \tag{2.11}$$

We have used two transformations on the non-stationary time series into the sequence of independent random variables $\mathbf{z}_k$ $(k=1,\cdots,n)$ which follows $N_p(\mathbf{0},\mathbf{\Sigma}_x + a_{kn}^{*}\mathbf{\Sigma}_v)$, and the coefficients a_{kn}^{*} is a dense sample of $4\sin^2(x)$ in $(0,\pi/2)$ [1].

It may be natural to use $\mathbf{z}_k\mathbf{z}_k^{'}$ to estimate $a_{kn}^{*}\mathbf{\Sigma}_v + \mathbf{\Sigma}_x$ since it is the variance-covariance matrix of $\mathbf{z}_k$. We notice that $a_{kn}^{*} \to 0$ as $n \to \infty$ for a fixed k. When k is small, a_{kn}^{*} is small and we can expect that $k = k_n$ depending n is still small when n is large. However, $(1/m_n)\sum_{k=1}^{m_n} a_{kn}^{*}$ is not small if m_n is near to n, which suggests the condition $m_n/n \to 0$ as $n \to \infty$. The separating information maximum likelihood (SIML) estimator of $\hat{\mathbf{\Sigma}}_x$ can be defined by

$$\hat{\mathbf{\Sigma}}_{x,SIML} = \frac{1}{m_n}\sum_{k=1}^{m_n} \mathbf{z}_k\mathbf{z}_k^{'} \ . \tag{2.12}$$

This estimator of the variance-covariance of non-stationary trends is trying to use the information on trends in the frequency domain, which corresponds to only the trend parts without measurement errors from the time series observations. For $\hat{\mathbf{\Sigma}}_x$, the number of terms m_n should be dependent on n. Then we need the order requirement that $m_n = O(n^{\alpha})$ and $0 < \alpha < 1$, which is the first property of the macro-SIML estimation.

From our construction of the SIML estimation the essential features of estimation do not depend on the noise terms as long as the noise terms are stationary. It has been pointed out by Kunitomo and Sato (2017).

[1] We have used a_{kn}^{*}, which is slightly different a_k in Kunitomo and Sato (2013) and the latter corresponds to $a_{kn} = na_{kn}^{*}$.

3. Simple Cases

3.1 An Illustration

Let y_i be the $i-$th observation of time series for $i = 1, \cdots, n$ and $\mathbf{y}_n = (y_i)$ be an $n \times 1$ vector of observations. (y_0 is the initial observation.) We consider the situation when the underlying non-stationary trends x_i at $i = 1, \cdots, n$ are not necessarily the same as the observed time series and followed by

$$x_i = x_{i-1} + v_i^{(x)} \tag{3.1}$$

where $v_i^{(x)}$ are the independent random variables followed by $N(0, \sigma_x^2)$ and x_0 is the initial variable. Let v_i are the independent random variables followed by $N(0, \sigma_v^2)$, which are independent of x_i ($i = 1, \cdots, n$). Then we consider the additive model

$$y_i = x_i + v_i \ \ (i = 1, \cdots, n) \ . \tag{3.2}$$

Then the log-likelihood function is proportional to

$$L_n = \sum_{k=1}^{n} \log |a_{kn}^{*} \sigma_v^2 + \sigma_x^2|^{-1/2} - \frac{1}{2} \sum_{k=1}^{n} \frac{z_k^2}{a_{kn}^{*} \sigma_v^2 + \sigma_x^2} \ , \tag{3.3}$$

where

$$a_{kn}^{*} = 4 \sin^2 \left[\frac{\pi}{2} \left(\frac{2k-1}{2n+1} \right) \right] \ (k = 1, \cdots, n) \ . \tag{3.4}$$

By defining $\tau = \sigma_x^2 / \sigma_v^2$ (≥ 0), we rewrite $-(1/2) L_n$ as

$$L_{1n} = \sum_{k=1}^{n} [\log \sigma_v^2 + \log(a_{kn}^{*} + \tau)] + \frac{1}{\sigma_v^2} \sum_{k=1}^{n} \frac{z_k^2}{a_{kn}^{*} + \tau} \ . \tag{3.5}$$

Since $z_k \sim N(0, a_{kn}^{*} \sigma_v^2 + \sigma_x^2)$ $(k = 1, \cdots, n)$, then the maximum likelihood estimator of σ_v^2 is given by

$$\hat{\sigma}_{v.ML}^2 = \frac{1}{n} \sum_{k=1}^{n} \frac{z_k^2}{a_{kn}^{*} + \tau} \tag{3.6}$$

and the concentrated (normalized) likelihood function) in this simple case is proportional to

$$L_{1n}(\tau) = \log[\frac{1}{n} \sum_{k=1}^{n} \frac{z_k^2}{a_{kn}^{*} + \tau}] + \frac{1}{n} \sum_{k=1}^{n} \log[a_{kn}^{*} + \tau] + 1 \ . \tag{3.7}$$

It may not be straightforward to obtain the maximum likelihood estimator of τ because the likelihood function may not be a simple function. Also Hirotsugu Akaike (1985) argued that we should impose the restriction $0 < \tau \leq 1$ in our setting because of the usefulness of models. He advocated to estimate the models by using the ABIC maximization procedure.
On the other hand, we approximate (1.9) as

$$L_n^{SI} = \sum_{k=1}^{m_n} \log |a_{kn}^* \sigma_v^2 + \sigma_x^2|^{-1/2} - \frac{1}{2} \sum_{k=1}^{m_n} \frac{z_k^2}{a_{kn}^* \sigma_v^2 + \sigma_x^2} , \tag{3.8}$$

and we have some requirement on $m_n/n = o(1)$. We notice that $a_{kn}^* = o(1)$ when $k = 1, \cdots, m_n$ and $n \to \infty$. Then the SIML estimator σ_x^2 can be given by

$$\hat{\sigma}_{x.SIML}^2 = \frac{1}{m_n} \sum_{k=1}^{m_n} z_k^2 . \tag{3.9}$$

However, the macro-SIML estimation of σ_v^2 is not the same as the original (finance) SIML estimation because $a_{kn}^* = O(1)$ for $k = n - m_n + 1, \cdots, n$. One way to estimate σ_v^2 is to use the fact that

$$\mathcal{E}[\frac{1}{n} \sum_{i=1}^{n} z_k^2] = \sigma_x^2 + (\frac{1}{n} \sum_{i=1}^{n} a_{kn}^*) \sigma_v^2 \longrightarrow \sigma_x^2 + 2\sigma_v^2 \ (n \to \infty). \tag{3.10}$$

It is because $a_{kn}^* \longrightarrow 0$ when $k_n/n \to 0$ $(n \to \infty)$ and $a_{kn}^* \to 4$ $(k_n/n \to 1, n \to \infty)$. Then a possible SIML estimator σ_v^2 can be given by

$$\hat{\sigma}_{v.SIML}^2(1) = \frac{1}{2} [\frac{1}{n} \sum_{k=1}^{n} z_k^2 - \hat{\sigma}_{x.SIML}^2] , \tag{3.11}$$

with the restriction of the resulting positivity.
For the estimation problem of high-frequency financial data, Kunitomo and Sato (2013) have suggested to use

$$\hat{\sigma}_{v.SIML}^2 = \frac{1}{l_n} \sum_{k=n-l_n}^{n} a_{kn}^{-1} z_k^2 , \tag{3.12}$$

where $a_{kn} = n a_{kn}^*$ and $l_n = o(n)$ for the hight frequency asymptotics. However, in the present case of Macro-SIML with a_{kn}^* it is straightforward to show that $\frac{1}{l_n} \sum_{k=n-l_n+1}^{n} a_{kn}^{*-1} \to 1/4$ as $n \to \infty$, the macro-SIML estimator is given by

$$\hat{\sigma}_{v.SIML}^2(2) = \frac{1}{l_n} \sum_{k=n-l_n+1}^{n} a_{kn}^{*-1} z_k^2 - \frac{1}{4} \hat{\sigma}_{x.SIML}^2 \tag{3.13}$$

with the restriction of the resulting positivity.

3.2 A Non-stationary Common Trend Case

Let $\mathbf{y}_i$ be the $i-$th observation of p-dimensional time series for $i = 1, \cdots, n$ and $\mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i$. Also let $\mathbf{Y}_n = (\mathbf{y}_i^{'})$ be an $n \times p$ $(p \geq 1)$ matrix of observations. We assume that the vectors $\mathbf{x}_i$ satisfy

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \boldsymbol{\pi}\mu_i^* \ , \tag{3.14}$$

where $\boldsymbol{\pi}$ is a $p \times 1$ vector, μ_i^* is a sequence of i.i.d. (one-dimensional) random variables [2] following $N(0, \sigma_\mu^2)$ and $\mathbf{v}_i$ are i.i.d. (p-dimensional) random variables following $N_p(\mathbf{0}, \boldsymbol{\Sigma}_v)$ with the variance-covariance matrix $\boldsymbol{\Sigma}_v$, which is a non-singular matrix. We take $\boldsymbol{b} = \sigma_\mu \boldsymbol{\pi}, \mathbf{A} = a_{kn}^* \boldsymbol{\Sigma}_v$ and apply the matrix formulas such that for a positive definite $\mathbf{A}$ we have

$$|\mathbf{A} + \mathbf{b}\mathbf{b}^{'}| = |\mathbf{A}|[1 + \mathbf{b}^{'}\mathbf{A}^{-1}\mathbf{b}] \tag{3.15}$$

and

$$[\mathbf{A} + \mathbf{b}\mathbf{b}^{'}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{b}[1 + \mathbf{b}^{'}\mathbf{A}^{-1}\mathbf{b}]^{-1}\mathbf{b}^{'}\mathbf{A}^{-1} \tag{3.16}$$

for $\boldsymbol{\Sigma}_x = \mathbf{b}\mathbf{b}^{'}$.
Then the likelihood function L_n is proportional to $(-1/2)$ times

$$\begin{aligned} L_{1n} &= \sum_{k=1}^{n} \left[\log |a_{kn}^* \boldsymbol{\Sigma}_v| + \log(1 + a_{kn}^{*-1}\mathbf{b}^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b}) + a_{kn}^{*-1}\mathbf{z}_k^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{z}_k - \frac{a_{kn}^{*-1}(\mathbf{z}_k^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b})^2}{a_{kn}^* + \mathbf{b}^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b}} \right] \\ &= \sum_{k=1}^{n} \log |a_{kn}^* \boldsymbol{\Sigma}_v| + \sum_{k=1}^{n} a_{kn}^{*-1}\mathbf{z}_k^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{z}_k + \sum_{k=1}^{n} \left[\log(1 + a_{kn}^{*-1}c) - \frac{a_{kn}^{*-1}(\mathbf{z}_k^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b})^2}{a_{kn}^* + c} \right] \ , \end{aligned}$$

where we set

$$c = \sigma_\mu^2 \boldsymbol{\pi}^{'}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\pi} \ . \tag{3.17}$$

We need a normalization for the vector $\boldsymbol{\pi}$. If we take a simple normalization, the maximum likelihood estimator of $\boldsymbol{\pi}$ could be a quite complicated solution of the likelihood equation even when $p = 2$. One possible normalization is to take $\boldsymbol{\beta}^{'} = (1, -\boldsymbol{\beta}_2^{'})$ and then the maximization is not a trivial task.

As an alternative way to solve the present problem is to use the conditions that

$$\mathcal{E}[\mathbf{z}_k\mathbf{z}_k^{'}] = \boldsymbol{\Sigma}_x + o(1) \ \ \text{for} k = 1, \cdots, m_n$$

and

$$\mathcal{E}[a_{kn}^{*-1}\mathbf{z}_k\mathbf{z}_k^{'}] = \boldsymbol{\Sigma}_v + \frac{1}{4}\boldsymbol{\Sigma}_x + o(1) \ \ \text{for} k = n + 1 - m_n, \cdots, n \ .$$

[2]The notation μ_i^* is different from μ_i and $\mu_i^* = \Delta\mu_i$ $(i = 2, \cdots, n)$ in Kunitomo and Sato (2017).

The rank of matrix $\mathbf{\Sigma}_x$ is one while the matrix $\mathbf{\Sigma}_v$ has a full rank. Then it may be natural to consider the characteristic equation

$$\left[\hat{\mathbf{\Sigma}}_{x.SIML} - \lambda\hat{\mathbf{\Sigma}}_{v.SIML}\right]\hat{\boldsymbol{\beta}}_{SIML} = \mathbf{0}\ , \tag{3.18}$$

and

$$\hat{\Sigma}_{x.SIML} = \frac{1}{m_n}\sum_{k=1}^{m_n}\mathbf{z}_k\mathbf{z}_k^{'}\ , \tag{3.19}$$

$$\hat{\Sigma}_{v.SIML}(1) = \frac{1}{2}[\frac{1}{n}\sum_{k=1}^{n}\mathbf{z}_k\mathbf{z}_k^{'} - \hat{\Sigma}_{x.SIML}]\ , \tag{3.20}$$

or

$$\hat{\Sigma}_{v.SIML}(2) = \frac{1}{l_n}\sum_{k=n+1-l_n}^{n} a_{kn}^{*-1}\mathbf{z}_k\mathbf{z}_k^{'} - \frac{1}{4}\hat{\Sigma}_{x.SIML}\ , \tag{3.21}$$

where we have the restrictions of the resulting positive definiteness of the estimated variance-covariance matrix.
We use the notation that

$$\mathbf{Z}_n = (\mathbf{z}_k^{'}) = \mathbf{P}_n\mathbf{C}_n^{-1}\left(\mathbf{Y}_n - \mathbf{1}_n\bar{\mathbf{y}}_0^{'}\right)\ , \tag{3.22}$$

$\hat{\mathbf{\Sigma}}_{v.SIML}$ is a SIML estimator of $\mathbf{\Sigma}_v$, and λ is the (scalar) eigen value. Because the rank of $\mathbf{\Sigma}_x$ is degenerated and it is one, we need to take the smallest eigenvalue λ_1. We have the $\hat{\boldsymbol{\beta}}_{SIML}$, which is called the SIML estimator of $\boldsymbol{\beta}$. A simplified (consistent) estimation may be given by

$$\hat{\mathbf{\Sigma}}_{x.SIML} \times \hat{\boldsymbol{\beta}}_{SIL} = \mathbf{0}\ , \tag{3.23}$$

that is,

$$\hat{\mathbf{\Sigma}}_{x.SIML} \times [\begin{array}{c} 1 \\ -\hat{\boldsymbol{\beta}}_{2.SIL} \end{array}] = \mathbf{0}\ . \tag{3.24}$$

We can solve as

$$\hat{\boldsymbol{\beta}}_{2.SIL} = \hat{\mathbf{\Sigma}}_{22x.SIML}^{-1}\hat{\mathbf{\Sigma}}_{21x.SIML}\ , \tag{3.25}$$

which can be the least squares estimator for the transformed variables and it is called the SILS estimator.

3.3 On Autocorrelations of Noise terms

The main statistical problem of the time series decomposition models in Kunitomo and Sato (2017) was the estimation of non-stationary trend components and seasonal components. The results do not much depend on the specification of noise components. When there are autocorrelations in the noise terms as (2.3), the analysis of frequency domain on the underlying errors-in-variables models may give some insight on the issue.

When the vector sequence of noise terms $\mathbf{v}_i$ $(i = 1, \cdots, n)$ follows the stationary stochastic process as (2.3), we can represent the spectral density $p \times p$ matrix of $\mathbf{v}_i$ as

$$f_v(\lambda) = \frac{1}{\pi}\Big(\sum_{j=-\infty}^{\infty} \mathbf{C}_j e^{2i\lambda j}\Big)\mathbf{\Sigma}_e\Big(\sum_{j=-\infty}^{\infty} \mathbf{C}_j e^{-2i\lambda j}\Big) \quad \Big(-\frac{\pi}{2} \le \lambda \le \frac{\pi}{2}\Big), \tag{3.26}$$

where $i^2 = -1$ (see Chapter 7 of Anderson (1971) for instance.)
Then the $p \times p$ spectral density matrix of the transformed vector process $\Delta\mathbf{y}_i$ $(= \mathbf{y}_i - \mathbf{y}_{i-1})$ can be also represented as

$$f_{\Delta y}(\lambda) = \frac{1}{\pi}\left[\mathbf{\Sigma}_x + (1 - e^{2i\lambda}) f_v(\lambda)(1 - e^{-2i\lambda})\right] \quad . \tag{3.27}$$

Since the transformed random vectors $\boldsymbol{z}_k$ $(k = 1, \cdots, n)$ correspond to the Fourier transformed vectors of $\Delta\mathbf{y}_i$ $(i = 1, \cdots, n)$ except the initial condition $\mathbf{y}_0$, it is possible to estimate the spectral density matrix $f_v(\lambda)$ from the sequence of observations of $\mathbf{z}_k$ $(k = 1, \cdots, n)$ in principle.

The efficient estimation on $f_v(\lambda)$ could be certainly possible, but we need to investigate further related problems such as the kernel estimation and the effects of small sample sizes usually encounterd in analysing macro-economic time series by the multivariate economic time series, which are quite different from the initial purrpose of the present investigation,it is beyond scope of this paper and we will take have another occasion.

4. Gaussian Likelihood Function and Maximum Likelihood Estimation

It may be natural to think that we could apply the general principle of the maximum likelihood (ML) method. For the resulting simplicity, we use the case of i.i.d. noises for the ML estimation. One of interesting aspects of the present problem is the fact the ML method does not necessary give a satisfactory solution. We illustrate this problem by using Example 1 in Section 2 of Kunitomo and Sato (2017). We set the true parameter values as $\sigma_\mu^2 = 0.4, \beta_2 = 1.0$ and

$$\mathbf{\Sigma}_v = \begin{pmatrix} 0.45 & 0.23 \\ 0.23 & 0.4 \end{pmatrix} , \ \mathbf{\Sigma}_x = \sigma_\mu^2 \boldsymbol{\pi}\boldsymbol{\pi}' \ , \ \boldsymbol{\pi} = \begin{pmatrix} 1 \\ -\beta_2 \end{pmatrix} .$$

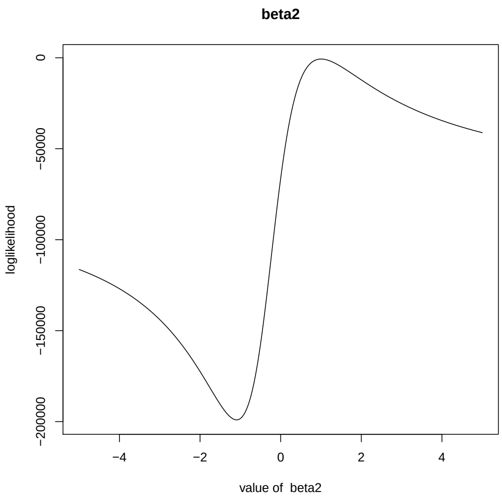

Figure 1 : Likelihood Function of β_2 $(n = 1,000)$

Then we generate a set of simulated observations as a typical example and we have drawn two Gaussian likelihood functions of β_2 in Figures 1-3 when $1,000$, given the true values for other parameters. We have found that the likelihood function could have some peculiar forms in some cases as illustrated by Figures 1 and 2. Also we have found that the Gaussian likelihood function is rather flat with respect to the correlation coefficient parameter. It seems that these are some of important consequences in the non-stationary errors-in-variables models.

We have drawn one wrong likelihood function in Figure 4 as an illustration on the assumption of Gaussian distributions. We generated the random variables followed by the uniform distribution on $[-2, +2]$. We have found that the variance-covariance estimation crucially depends on the assumption of Gaussianity as we had expected.

Now we investigate the analytical properties of the likelihood function in a detail.

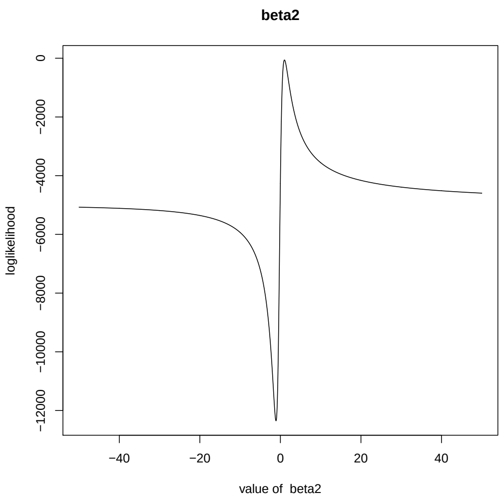

Figure 2 : Likelihood Function of β_2 $(n = 1,000)$

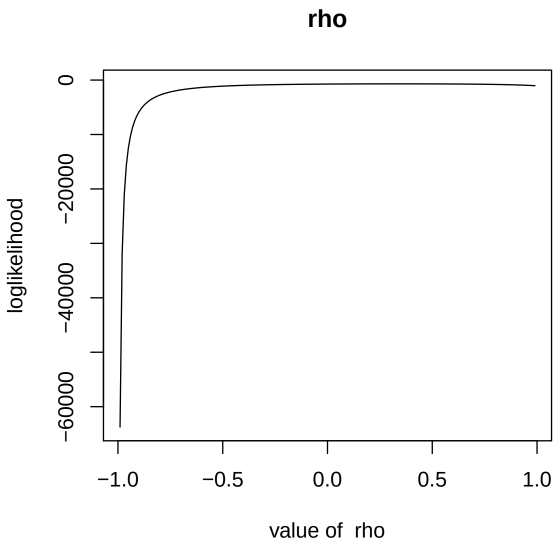

Figure 3 : Likelihood Function of ρ $(n = 1,000)$

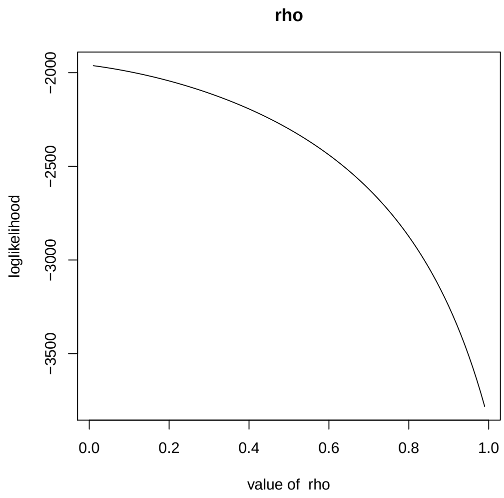

Figure 4 : Wrong Likelihood Function of ρ $(n = 1,000)$

We normalize the Gaussian likelihood function by n and rewrite

$$\begin{aligned}
L_{1n}^* &= \frac{1}{n}\sum_{k=1}^{n}|a_{kn}^*\mathbf{\Sigma}_v| + \frac{1}{n}\sum_{k=1}^{n}\log(1+a_{kn}^{*-1}c) \\
&\quad + \frac{1}{n}\sum_{k=1}^{n} a_{kn}^{*-1}\mathbf{tr}\left[\mathbf{\Sigma}_v^{-1}(\mathbf{\Sigma}_v - \frac{1}{a_{kn}^*+c}\mathbf{b}\mathbf{b}^{'})\mathbf{\Sigma}_v^{-1}(\mathbf{z}_k\mathbf{z}_k^{'} - (a_{kn}^*\mathbf{\Sigma}_v(\theta_0)+\mathbf{b}(\theta_0)\mathbf{b}(\theta_0)^{'}))\right] \\
&\quad + \frac{1}{n}\sum_{k=1}^{n} a_{kn}^{*-1}\mathbf{tr}\left[\mathbf{\Sigma}_v^{-1}(\mathbf{\Sigma}_v - \frac{1}{a_{kn}+c}\mathbf{b}\mathbf{b}^{'})\mathbf{\Sigma}_v^{-1}(a_{kn}^*\mathbf{\Sigma}_v(\theta_0)+\mathbf{b}(\theta_0)\mathbf{b}(\theta_0)^{'}))\right] \\
&= \frac{1}{n}\sum_{k=1}^{n}|a_{kn}^*\mathbf{\Sigma}_v| + L_{12n}^* + L_{13n}^* + L_{14n}^* \;\; (say),
\end{aligned}$$

where $\mathbf{\Sigma}_v(\theta_0)$ andf $\mathbf{b}(\theta_0)$ are evaluated at the true parameter values.
We prepare the next lemma.

Lemma 4.1 : Let a $p \times p$ random vector $\mathbf{z}_k$ follows $N_p(\mathbf{0},\mathbf{Q})$. Then for any $p \times p$ matric $\mathbf{A}_k$,

$$\mathcal{E}[(\mathrm{tr}(\mathbf{A}_k\mathbf{z}_k\mathbf{z}_k^{'}))^2] = [\mathrm{tr}(\mathbf{A}_k\mathbf{Q})]^2 + 2\mathrm{tr}(\mathbf{A}_k\mathbf{Q}\mathbf{A}_k\mathbf{Q}) \;. \tag{4.1}$$

By using this lemma, as $n \longrightarrow \infty$ the second term converges to

$$L_{13n}^* \xrightarrow{p} 0 \;. \tag{4.2}$$

Then we can establish the next result and the proof is given in Appendix. Although we expect that the ML estimation under the Gaussian assumption has good asymptotic properties, we could not find any proof in the present setting and then decided to give the proof. We can also expect the asymptotic normality of the ML estimator, but we have omitted its discussion because we need more spaces.

Theorem 4.2 : Assume that $\mathbf{v}_i$ $(i = 1, \cdots, n)$ are a squence of i.i.d. random vectors and $|\boldsymbol{\Sigma}_v| \neq 0$. Then under the assumption of Gaussian distributions the maximum likelihood estimator of $\boldsymbol{\beta}$ is consistent as $n \longrightarrow \infty$.

Remark 4.1 : It should be noted that in time series econometrics it has been known that coefficient parameter vector $\boldsymbol{\beta}$ can be estimated by using the standard regression method if the observed variables are co-integrated. However, the ML method and the SIML method estimate not only the coefficient vector consistently, but also the variance-covariance matrices of trend terms and the noise terms at the same time. Johansen (1995) had developed the ML method without any noise terms as well as seasonality. His main interested was the hypothesis testing problem of the rank condition.

We should mention to the first part of Theorem 4.1 of Kunitomo and Sato (2017) on the asymptotic property of the SIML estimator under general conditions. They have given the asymptotic normality of the SIML estimator under general moment conditions.

Theorem 4.3 (Kunitomo and Sato 2017) : Assume the non-stationary errors-in-variables model given by (2.1)-(2.3) and $|\boldsymbol{\Sigma}_v| \neq 0$. Then under the assumption of existence of fourth order moments the SIML estimator of $\boldsymbol{\beta}$ is consistent as $n \longrightarrow \infty$.

Although the SIML estimator has asymptotic normality, we have omitted this aspect in this paper. (See Kunitomo and Sato (2017) for the details.)

5. Simulations

There has not been any simulation result on alternative estimation methods for the non-stationary time series models with errors-in-variables as far as we know. It may be because we need a careful treatment of the non-stationarity with errors-in-variables. However, since there are many situations with macro-economics variables that we observe the non-stationarity and measurement errors, it is worthwhile to investigate the related issues by using simulations.

In our simulations of the SIML method, we have set $\sigma_\mu^2 = 1, \sigma_v^2 = 0.5, 2$ or 4, $\beta_2 = 1.5$, $n = 80$ or 400 and $m_n = n^\alpha$ with $\alpha = 0.6$ or 0.7 as the parametrization, and the number of Monte Carlo repetition is 1,500 in each case. We can summarize our setting of simulations as

$$\Sigma_x = \begin{pmatrix} \Sigma_{x,11} & \Sigma_{x,12} \\ \Sigma_{x,12} & \Sigma_{x,22} \end{pmatrix} = \begin{pmatrix} 2.25 & 1.5 \\ 1.5 & 1 \end{pmatrix}, \ \Sigma_v = \begin{pmatrix} \Sigma_{v,11} & \Sigma_{v,12} \\ \Sigma_{v,12} & \Sigma_{v,22} \end{pmatrix} = \begin{pmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}.$$

Among many simulations we have summarize them as Table 1. In Table 1 the number inside the parentheses are the standard deviation of estimators calculated by our simulations.

σ_v^2	α	n	$\Sigma_{x,11}$	$\Sigma_{x,12}$	$\Sigma_{x,22}$	$\Sigma_{v,11}(2)$	$\Sigma_{v,12}(2)$	$\Sigma_{v,22}(2)$	$\beta_{2.SIL}$	$\beta_{2.SIML}(2)$
0.5	0.6	80	2.263	1.475	1.038	0.495	0.010	0.499	1.420	1.528
			(0.904)	(0.602)	(0.415)	(0.445)	(0.305)	(0.318)	(0.114)	(0.178)
		400	2.239	1.484	1.006	0.512	0.009	0.501	1.475	1.502
			(0.542)	(0.362)	(0.244)	(0.283)	(0.185)	(0.186)	(0.038)	(0.038)
	0.7	80	2.294	1.454	1.086	0.521	0.030	0.496	1.339	1.554
			(0.706)	(0.469)	(0.337)	(0.386)	(0.250)	(0.251)	(0.131)	(0.269)
		400	2.296	1.499	1.044	0.498	0.007	0.494	1.436	1.502
			(0.393)	(0.263)	(0.182)	(0.214)	(0.137)	(0.136)	(0.047)	(0.054)
2	0.6	80	2.378	1.438	1.163	1.992	0.006	1.922	1.233	1.630
			(0.948)	(0.617)	(0.455)	(1.020)	(0.699)	(0.852)	(0.231)	(0.810)
		400	2.318	1.500	1.060	1.981	0.006	1.991	1.415	1.504
			(0.534)	(0.352)	(0.245)	(0.615)	(0.411)	(0.535)	(0.077)	(0.082)
	0.7	80	2.629	1.452	1.438	1.943	0.032	1.944	1.017	1.701
			(0.833)	(0.543)	(0.471)	(0.816)	(0.570)	(0.709)	(0.230)	(0.901)
		400	2.410	1.479	1.166	1.975	0.007	1.946	1.267	1.514
			(0.534)	(0.440)	(0.287)	(0.465)	(0.309)	(0.407)	(0.093)	(0.123)
4	0.6	80	2.639	1.469	1.384	3.927	-0.010	3.990	1.072	1.702
			(1.040)	(0.669)	(0.547)	(1.837)	(1.214)	(1.670)	(0.296)	(1.042)
		400	2.377	1.503	1.127	3.933	-0.008	3.965	1.334	1.514
			(0.558)	(0.368)	(0.267)	(1.099)	(0.727)	(1.025)	(0.105)	(0.132)
	0.7	80	3.118	1.457	1.885	3.777	0.065	3.874	0.787	1.846
			(1.005)	(0.636)	(0.630)	(1.427)	(0.963)	(1.311)	(0.274)	(1.452)
		400	2.601	1.483	1.357	3.937	0.010	3.908	1.095	1.519
			(0.451)	(0.298)	(0.249)	(0.806)	(0.550)	(0.727)	(0.119)	(0.197)

Table 1: Sample mean of estimators

We have done a number of simulations, but the results are similar to Table 1. There are several interesting findings. First, on the effects of sample sizes the performance of the estimators of the SIML estimation becomes as the sample size increases as we had expected. Second, when the variances of noises are small, both the SILS estimator and the SIML estimator give reasonable estimates on the coefficient parameter, the former is slightly biased toward zero while the latter has some correction of this bias. Third, when when the variances of noises are not small, the SILS estimator has a significant bias.

In addition to these simulations we have done a number of simulations when the underlying distributions are not true and there are some stationary noises instead of i.i.d. noises. In order to investigate the effects of autocorrelations in the noise terms, we consider the $VAR_2(1)$ model given by

$$\begin{aligned} v_i &= \Phi v_{i-1} + \bar{v}_i, \\ v_1 &\sim N(0, \Sigma_v),\ \bar{v}_i \overset{i.i.d.}{\sim} N(0, \bar{\Sigma}_v), \\ vec(\bar{\Sigma}_v) &= (I_4 - \Phi \otimes \Phi) vec(\Sigma_v), \\ \Phi &= diag(0.5, 0.5),\ \Sigma_v = \begin{pmatrix} 0.4 & 0.3\sqrt{0.4 \times 0.45} \\ 0.3\sqrt{0.4 \times 0.45} & 0.45 \end{pmatrix} \end{aligned}$$

Then we summarize the simulation results as Table 2. Inside the parentheses in Table 2 are standard deviation of estimators.

	n	$\sigma^2_{x.SIML}$	$\beta_{2.SIL}$	$\beta_{2.SIML}(2)$
$\beta_2 = -1.5$	100	0.234 (0.083)	-1.154 (0.275)	-1.512 (0.469)
	200	0.225 (0.065)	-1.208 (0.176)	-1.493 (0.330)
	400	0.216 (0.045)	-1.281 (0.112)	-1.497 (0.175)
$\beta_2 = -0.5$	100	0.251 (0.086)	-0.258 (0.161)	-0.456 (0.467)
	200	0.221 (0.052)	-0.337 (0.102)	-0.488 (0.226)
	400	0.208 (0.048)	-0.403 (0.074)	-0.494 (0.121)
$\beta_2 = 0.5$	100	0.264 (0.105)	0.436 (0.115)	0.601 (0.345)
	200	0.224 (0.061)	0.465 (0.102)	0.572 (0.242)
	400	0.214 (0.042)	0.498 (0.041)	0.552 (0.113)
$\beta_2 = 1.5$	100	0.232 (0.144)	1.183 (0.196)	1.714 (0.773)
	200	0.209 (0.076)	1.274 (0.162)	1.585 (0.507)
	400	0.204 (0.054)	1.341 (0.079)	1.509 (0.147)

Table 2: Sample mean of estimators

As we had expected, the SIML estimation of the non-stationary part does not depend on the autocorrelation structure of noise terms while the ML estimation depends on the true structure of the underlying process. Although both the ML estimation and the SIML estimation give similar stable results on the coefficient β_2 in some cases, the ML estimation has sometimes computational difficulties when the absolute size of coefficients are large. (It is the case when β_2 is relatively large as $|\beta_2| = 1.5$ in our simulations.) It may be due to the shape of the likelihood function becomes some peculiar shapes in such cases [3]. But we have omitted the details of a large number of simulations.

To summarize our simulations, the finite sample performance of the SIML estimation gives reasonable performances as the asymptotic theory has suggested as in the previous sections.

6. Extensions

There can be several extensions of the problem we have been investigating and the results obtained by Kunitomno and Sato (2017).

First, for the multivariate non-stationary (economic) time series, there are possibilities of co-integration in trends. In our framework, it may be interesting to consider the general case of reduced rank cases when

$$\text{rank}[\mathbf{\Sigma}_x] = q \ \ , \ \ 1 \leq q \leq p \ , \tag{6.1}$$

where $\mathbf{\Sigma}_x = \mathbf{B}\mathbf{B}^{'}$ and $\mathbf{B}$ is a $p \times q$ matrix.
Then the first example of Section 2 corresponds to the case when $p = q = 1$ and the second example of Section 3 corresponds to the case when $q = 1 < p, p \geq 2$.
In the general case, however, there is a parametrization problem for the $p \times p$ matrix $\mathbf{\Sigma}_x$, whose rank is q. We take a normalization as $\mathbf{B}^{'}\mathbf{\Sigma}_v^{-1}\mathbf{B} = (\text{diag } c_{ii})$. Then the algebra of Section 3.2 can be extended by using the matrix formulae such that for a positive definite $\mathbf{A}$ we have

$$|\mathbf{A} + \mathbf{B}\mathbf{B}^{'}| = |\mathbf{A}||\mathbf{I}_q + \mathbf{B}^{'}\mathbf{A}^{-1}\mathbf{B}]| \tag{6.2}$$

and

$$[\mathbf{A} + \mathbf{B}\mathbf{B}^{'}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}[\mathbf{I}_q + \mathbf{B}^{'}\mathbf{A}^{-1}\mathbf{B}]^{-1}\mathbf{B}^{'}\mathbf{A}^{-1} \tag{6.3}$$

for $\mathbf{\Sigma}_x = \mathbf{B}\mathbf{B}^{'}$.

Theorem 6.1 : Assume $1 \leq q < p, p \geq 2$ and $|\mathbf{\Sigma}_v| \neq 0$.
(i) Assume that $\mathbf{v}_i$ $(i = 1, \cdots, n)$ are a squence of i.i.d. random vectors and $|\mathbf{\Sigma}_v| \neq 0$. Under the assumptions of Gaussian distributions, the ML estimator of $\mathbf{B}$ is consistent as $n \longrightarrow \infty$.

[3] We are investigating this computational problem of the ML solutions in non-stationary multi-dimension cases.

(ii) Assume the non-stationary errors-in-variables model given by (2.1)-(2.3) and $|\mathbf{\Sigma}_v| \neq 0$. Under the assumption of existence of fourth order moments, the SIML estimator of $\mathbf{B}$ is consistent as $n \longrightarrow \infty$.

Second, in some cases the second order (or higher order) differencing may be often appropriate for economic time series. Since the likelihood function can be complicated in general, we consider the ML estimation and the SIML estimation when $p \geq 1$ and $d = 2$, where

$$\Delta^d \mathbf{x}_i = \mathbf{v}_i^{(x)} , \tag{6.4}$$

$\mathcal{E}[\mathbf{v}_i^{(x)}] = \mathbf{0}$, and $\mathcal{E}[\mathbf{v}_i^{(x)} \mathbf{v}_i^{(x)'}] = \mathbf{\Sigma}_x$.
We use the $\mathbf{K}_n-$transformation that from the observation matrix $\mathbf{Y}_n$ to $\mathbf{Z}_n^{(2)}$ $(= (\mathbf{z}_k^{(2)'}))$ by

$$\mathbf{Z}_n^{(2)} = (\mathbf{z}_k^{(2)'}) = \mathbf{K}_n \left(\mathbf{Y}_n - \bar{\mathbf{Y}}_0 \right) \ , \mathbf{K}_n = \mathbf{P}_n \mathbf{C}_n^{-2} \ . \tag{6.5}$$

Then the separating information maximum likelihood (SIML) estimator of $\hat{\mathbf{\Sigma}}_x$ in this case can be defined by

$$\hat{\mathbf{\Sigma}}_{x,SIML} = \frac{1}{m_n} \sum_{k=1}^{m_n} \mathbf{z}_k^{(2)} \mathbf{z}_k^{(2)'} \ . \tag{6.6}$$

We prepare the next Lemma.

Lemma 6.2 : Let

$$\mathbf{K}_n = (b_{ij}^{(2)}) = \mathbf{P}_n \mathbf{C}_n^{-2} \ . \tag{6.7}$$

Then for $i, i' = 1, \cdots, m_n$, we have

$$\sum_{j=1}^{n} b_{ij}^{(2)} b_{i',j}^{(2)} = \delta(i, i') \left[2 \sin(\frac{\pi}{2} \frac{2i-1}{2n+1}) \right]^4 + O(\frac{1}{n}) \ . \tag{6.8}$$

By using the above lemma, we have the next result, which is an extension of Kunitomo and Sato (2017) for the case of $d \geq 1$. The result holds for an arbitrary integer d.

Theorem 6.3 : Assume $p \geq 1$, $d = 2$ and $m_n/n \longrightarrow 0$ as $n \longrightarrow \infty$.
(i) Assume that $\mathbf{v}_i$ $(i = 1, \cdots, n)$ are a squence of i.i.d. random vectors and $|\mathbf{\Sigma}_v| \neq 0$. Under the assumption of Gaussian distributions, the ML estimator of $\mathbf{\Sigma}_x$ is consistent as $n \longrightarrow \infty$.
(ii) Assume the non-stationary errors-in-variables model given by (2.1)-(2.3) and

$|\mathbf{\Sigma}_v| \neq 0$. Under the assumption of existence of fourth order moments, the SIML estimator of $\mathbf{\Sigma}_x$ is consistent as $n \longrightarrow \infty$.

It should be important to note that the diagonal elements a_{kn} $(k = 1, \cdots, n)$ should be modified to

$$a_{kn}^{(2)} = [2 \sin \frac{\pi}{2n+1}(k - \frac{1}{2})]^4 \tag{6.9}$$

in the present case and we need the corresponding bias correction for estimating the variance-covariance matrix $\mathbf{\Sigma}_v$.

Remark 6.1 : The SIML should be a useful tool for the state space modeling of non-stationary multivariate time series because it does not have any computational problem and it has an asymptotic robustness. Also it is possible to show that the asymptotic normality of the SIML estimator under the general condition of (2.3) with additional arguments ouitlined in Kunitomo and Sato (2017). For instance, the SIML method may give reasonable estimates not only for the coefficients parameters, but also the variance-covariance matrices when $d = 1$ and $d = 2$ by taking m_n appropriately.

7. Concluding Remarks

In this study, we have reported the asymptotic properties and finite sample properties of two estimation methods for the non-stationary errors-in-variables models. We have compared the SIML estimation and the maximum likelihood (ML) estimation when there are non-stationary trends and noise components. We have found that the Gaussian likelihood function shows some peculiar shape and the computation of the ML estimation can be unstable in some cases. On the other hand, the SIML estimation has the asymptotic robust properties under general conditions of existence of second order moments. We have investigated the conditions for the good asymptotic properties and the finite sample properties of the ML estimation and SIML estimation by both the theoretical analysis and simulations. We have found that the SIML method gives reasonable estimates not only for the coefficients parameters, but also the variance-covariance matrices when $d = 1$ and $d = 2$ by taking m_n appropriately.

There are several possible extensions and discussions. First, it may be interesting to see to what extent the results reported in this paper are still valid when there are seasonality and stationary components although the issue has been investigated by Kunitomo and Sato (2017) in a simple setting. Second, there are several important issues on modeling the non-stationary trends and noises. Although we have discussed some of them briefly in Section 6, but we need more investigations. The determination of the number of non-stationary factors is one of examples, which is discussed by Kunitomo (2017). Third, there can be many applications of the methods we have discussed as indicated in Kunitomo and Sato (2017).

The results of current investigations on these issues will be reported in other occasions.

References

[1] Akaike, H. (1985), "Prediction and entropy. A Celebration of Statistics," A. C. Atkinson and E. Fienberg, eds., Springer-Verlag, New York, 1-24.

[2] Akaike, H. (1989), "Choosing priors and its applications," in *Bayesian Statistics and Applications*, edited by Suzuki and Kunitomo, Tokyo University Press, in Japanese.

[3] Amemiya, T. (1986), *Advanced Econometrics*, Blackwell.

[4] Anderson, T.W. (1971), *Statistical Time Series Analysis*, John-Wiley.

[5] Anderson, T.W. (1984), "Estimating Linear Statistical Relationships," *Annals of Statistics*, 12, 1-45.

[6] Anderson, T.W. (2003), *An Introduction to Statistical Multivariate Analysis*, 3rd Edition, John-Wiley.

[7] Chang, Y., B. Jiang and J. Park (2008), "Using Kalman Filter to Extract and Test for Common Stochastic Trends," Unpublished Manuscript.

[8] Fuller, W. (1987), *Measurement Error Models*, John-Wiley.

[9] Johansen, S. (1995), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, Oxford UP.

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

[11] Kunitomo, N. (2017), "On Estimating the Number of Factors in Non-stationary Errors-in-Variables Models," In preparation.

[12] Kunitomo, N. and S. Sato (2013), "Separating Information Maximum Likelihood Estimation of Realized Volatility and Covariance with Micro-Market Noise," *North American Journal of Economics and Finance* (2013), 26, 282-309.

[13] Kunitomo, N. (2017), "On Detecting the Number of Factors in Non-stationary Errors-in-Variables Models," Unpublished Manuscript.

[14] Kunitomo, N. and S. Sato (2017), "Trend, Seasonality and Economic Time Series : the Nonstationary Errors-in-variables Models," Discussion Paper CIRJE-F-977, Graduate School of Economics, University of Tokyo,(http://www.e.u-tokyo.ac.jp/cirje/research/dp/2015), revised.

[15] Ohtsu, K., H. Pend and G. Kitagawa, (2015), *Time Series Modeling for Analysis and Control Advanced Autopilot and Monitoring Systems*, Springer-Brief in Statistics, Springer.

APPENDIX : Mathematical Derivations

In this Appendix, we give some details of the derivations omitted in the previous sections.

Proof of Lemma 4.1 : When we have $\mathbf{z}_k = (z_{ik}) \sim N_p(\mathbf{0}, \mathbf{Q})$, we can use the relation

$$\mathcal{E}[z_{ik}z_{jk}z_{i'k}z_{j'k}] = q_{ij}q_{i'j'} + q_{ii'}q_{jj'} + q_{ij'}q_{i'j} \ , \tag{A.1}$$

where $\mathbf{Q} = (q_{ij})$. Then it is straightforward to obtain the results.
(**Q.E.D.**)

Proof of Theorem 4.2 : By using Lemma 4.1, it is possible to obtain the variance of L_{13n}^*, which converges to 0 as $n \longrightarrow \infty$. We set

$$\mathbf{A}_k = a_{kn}^{*-1}\mathbf{\Sigma}_v^{-1}[\mathbf{\Sigma}_v - \frac{1}{a_{kn}^* + c}\mathbf{b}\mathbf{b}^{'}]\mathbf{\Sigma}_v^{-1}$$

and

$$\mathbf{Q}_0 = a_{kn}^*\mathbf{\Sigma}_v(\theta_0) + \mathbf{b}(\theta_0)\mathbf{b}(\theta_0)^{'} \ .$$

Then

$$\mathrm{tr}[\mathbf{A}_k\mathbf{Q}_0] = \mathrm{tr}(a_{kn}^{*-1}\mathbf{\Sigma}_v^{-1}[\mathbf{\Sigma}_v - \frac{1}{a_{kn}^* + c}\mathbf{b}\mathbf{b}^{'}]\mathbf{\Sigma}_v^{-1}[a_{kn}\mathbf{\Sigma}_v(\theta_0) + \mathbf{b}(\theta_0)\mathbf{b}(\theta_0)^{'}]) \ .$$

If $\mathbf{Q} = \mathbf{Q}_0$, then we have

$$\mathrm{tr}(\mathbf{A}_k\mathbf{Q}_0) = \mathrm{tr}(\mathbf{I}_p) = p \ .$$

Also we find

$$\begin{aligned} (\mathrm{A.2}) \qquad & = \mathcal{E}[\sum_{k=1}^{n}\mathrm{tr}\mathbf{A}_k(\mathbf{z_k}\mathbf{z}_k^{'} - \mathbf{Q}_0)]^2 \\ & = \mathcal{E}[\sum_{k,k'=1}^{n}\mathrm{tr}\mathbf{A}_k(\mathbf{z}_k\mathbf{z}_k^{'} - \mathbf{Q}_0)\mathbf{A}_{k'}(\mathbf{z}_{k'}\mathbf{z}_{k'}^{'} - \mathbf{Q}_0)] \\ & = [\sum_{k=1}^{n}\mathcal{E}[(\mathbf{z}_k^{'}\mathbf{A}_k\mathbf{z}_k)^2 - (\mathrm{tr}\mathbf{A}_k\mathbf{Q}_0)^2] \\ & = \sum_{k=1}^{n}2\mathrm{tr}(\mathbf{A}_k\mathbf{Q}_0\mathbf{A}_k\mathbf{Q}_0) \ . \end{aligned}$$

If $\mathbf{Q} = \mathbf{Q}_0$, then it is $2pn$.
Next, we investigate the last two terms of L^*_{1n} in details. After simple abgebra, they can be reexpressed as

$$
\begin{aligned}
L^*_{14} &= \mathrm{tr}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}(\theta_0) + \frac{1}{n}\sum_{k=1}^{n} a_{kn}^{*-1}\mathbf{b}(\theta_0^{'})\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0) \\
&\quad -\frac{1}{n}\sum_{k=1}^{n}\frac{1}{a^*_{kn}+c}\mathbf{b}^{'}[\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}(\theta)\boldsymbol{\Sigma}_v^{-1} + a_{kn}^{*-1}\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0)\mathbf{b}(\theta_0)^{'}\boldsymbol{\Sigma}_v^{-1}]\mathbf{b} \\
&= \mathrm{tr}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}(\theta_0) \\
&\quad +\frac{1}{n}\sum_{k=1}^{n}\frac{1}{a^*_{kn}(a^*_{kn}+c)}\Big[(a^*_{kn}+c)[\mathbf{b}(\theta_0)^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0) - a_{kn}^{*-1}\mathbf{b}^{'}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}_v(\theta_0)\boldsymbol{\Sigma}_v^{-1}\mathbf{b} \\
&\quad -(\mathbf{b}^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta))^2\Big] \\
&= \mathrm{tr}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}(\theta_0) \\
&\quad +\frac{1}{n}\sum_{k=1}^{n}\frac{1}{a^*_{kn}(a^*_{kn}+c)}\Big[\mathbf{b}^{'}\boldsymbol{\Sigma}_v^{-1}(\boldsymbol{\Sigma}_v - \boldsymbol{\Sigma}_v(\theta_0))\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0) \\
&\quad +\mathbf{b}(\theta_0)^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0)\mathbf{b}^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b} - (\mathbf{b}^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0))^2\Big] \ .
\end{aligned}
$$

The last term is non-negative because of the Cauchy-Schwarz's inequality and its minimum occurs at $\mathbf{b} = \mathbf{b}(\theta_0)$ because $\boldsymbol{\Sigma}_v$ is positive definite. Then we need to evaluate the sum of non-degenerate terms as

$$
\text{(A.3)} \qquad \frac{1}{n}\sum_{k=1}^{n}\log|a^*_{kn}\boldsymbol{\Sigma}_v| + \frac{1}{n}\sum_{k=1}^{n}\log[1+a_{kn}^{*-1}\mathbf{b}^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b}]
$$

and

$$
\text{(A.4)} \qquad \mathrm{tr}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}_v(\theta_0) + \frac{1}{n}\sum_{k=1}^{n}\frac{\mathbf{b}^{'}\boldsymbol{\Sigma}_v^{-1}(\boldsymbol{\Sigma}_v - \boldsymbol{\Sigma}_v(\theta_0))\boldsymbol{\Sigma}_v^{-1}\mathbf{b}}{a^*_{kn}+\mathbf{b}^{'}\boldsymbol{\Sigma}_v^{-1}\mathbf{b}} \ .,
$$

By using the inversion formula and its determinant in (3.1) and (3.2), the sum of above two terms can be written as

$$
\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}\log|a^*_{kn}\boldsymbol{\Sigma}_v + \mathbf{b}^{'}\mathbf{b}| \\
+\ &\frac{1}{n}\sum_{k=1}^{n}\mathrm{tr}[a^*_{kn}\boldsymbol{\Sigma}_v + \mathbf{b}^{'}\mathbf{b}]^{-1}[a^*_{kn}\boldsymbol{\Sigma}_v(\theta_0) + \mathbf{b}^{'}\mathbf{b}] \ .
\end{aligned}
$$

By using Lemma 3.2.2 of Anderson (2003), the minimum of each terms occurs at

$$
a^*_{kn}\boldsymbol{\Sigma}_v + \mathbf{b}(\theta_0)\mathbf{b}(\theta_0)^{'} = a^*_{kn}\boldsymbol{\Sigma}_v(\theta_0) + \mathbf{b}(\theta_0)\mathbf{b}(\theta_0)^{'} \ ,
$$

that is, $\mathbf{\Sigma}_v = \mathbf{\Sigma}_v(\theta_0)$. Hence the global minimum of the likelihood function occurs iff $\mathbf{b} = \mathbf{b}(\theta_0)$ and $\mathbf{\Sigma}_v = \mathbf{\Sigma}_v(\theta_0)$.
The rest of our arguments for the consistency follows from the general arguments (see Theorem 4.1.1 of Amemiya (1985), for instance) and we have the result.
(**Q.E.D.**)

Proof of Lemma 6.2 : We set

$$\text{(A.5)} \qquad \theta_{kj} = \frac{2\pi}{2n+1}(j-\frac{1}{2})(k-\frac{1}{2}) \,,$$

$$\text{(A.6)} \qquad \theta_{k} = \frac{2\pi}{2n+1}(k-\frac{1}{2}) \,,$$

and

$$\text{(A.7)} \qquad b_{kj}^{(2)} = \frac{1}{\sqrt{2n+1}}[(1-e^{i\theta_k})^2 e^{i\theta_{kj}} + (1-e^{-i\theta_k})^2 e^{-i\theta_{kj}}]$$

for $j,k = 1,\cdots,n$. Then

$$\begin{aligned}
(2n+1)\sum_{j=1}^{n}[b_{kj}^{(2)}]^2 &= (1-e^{i\theta_k})^4\frac{1+e^{i\theta_k}}{1-e^{i2\theta_k}} + (1-e^{-i\theta_k})^4\frac{1+e^{-i\theta_k}}{1-e^{-i2\theta_k}} + 2n(1-e^{i\theta_k})^2(1-e^{-i\theta_k})^2 \\
&= (1-e^{i\theta_k})^3 + (1-e^{-i\theta_k})^3 + 2n(e^{i\frac{\theta_k}{2}}1 - e^{-i\frac{\theta_k}{2}})^4 \\
&= 2n[4\sin^2\frac{\theta_k}{2}]^2 + o(1) \,.
\end{aligned}$$

Also for $k \neq k^{'}$ we find that

$$\begin{aligned}
&(2n+1)\sum_{j=1}^{n}[b_{kj}^{(2)}b_{k^{'}j}^{(2)}] \\
= \;& (1-e^{i\theta_k})^2(1-e^{-i\theta_k})^2\sum_{j=1}^{n}e^{i(\theta_{kj}+\theta_{k^{'}j})} + (1-e^{i\theta_k})^2(1-e^{-i\theta_{k^{'}}})^2\sum_{j=1}^{n}e^{i(\theta_{kj}+\theta_{k^{'}j})} \\
& +(1-e^{-i\theta_k})^2(1-e^{i\theta_{k^{'}}})^2\sum_{j=1}^{n}e^{i(\theta_{kj}-\theta_{k^{'}j})} + (1-e^{-i\theta_k})^2(1-e^{-i\theta_{k^{'}}})^2\sum_{j=1}^{n}e^{-i(\theta_{kj}+\theta_{k^{'}j})} \\
= \;& (I)+(II)+(III)+(IV) \;\; (say) \,.
\end{aligned}$$

Then after some algebra, we have

$$\begin{aligned}
(I)+(IV) &= (-1)[(e^{-i\frac{\theta_k}{2}} - e^{i\frac{\theta_k}{2}})^2(e^{-i\frac{\theta_{k^{'}}}{2}} - e^{i\frac{\theta_{k^{'}}}{2}})^2]e^{i\frac{\theta_k+\theta_{k^{'}}}{2}} \,, \\
(II)+(III) &= (-1)[(e^{-i\frac{\theta_k}{2}} - e^{i\frac{\theta_k}{2}})^2(e^{-i\frac{\theta_{k^{'}}}{2}} - e^{i\frac{\theta_{k^{'}}}{2}})^2][1-(e^{-i\frac{(\theta_k-\theta_{k^{'}})}{2}} + (e^{i\frac{(\theta_k-\theta_{k^{'}})}{2}})] \,.
\end{aligned}$$

Hence we finnally find that

$$(I)+(II)+(III)+(IV) = (-1)(e^{-i\frac{\theta_k}{2}}-e^{i\frac{\theta_k}{2}})^2(e^{-i\frac{\theta_{k^{'}}}{2}}-e^{i\frac{\theta_{k^{'}}}{2}})^2[1-e^{i\frac{\theta_k-\theta_{k^{'}}}{2}}-e^{-i\frac{\theta_k-\theta_{k^{'}}}{2}}+e^{-i\frac{\theta_k-\theta_{k^{'}}}{2}}]$$

and then it is proportional to

$$(I) + (II) + (III) + (IV) = 4[\sin^2 \frac{\theta_k}{2}][\sin^2 \frac{\theta_{k'}}{2}] + o(1)$$

when $k/n \longrightarrow 0$ as $n \to \infty$.
(**Q.E.D.**)