ORIGINAL PAPER

Comparing estimation methods of non-stationary errors-in-variables models

Naoto Kunitomo[1] · Naoki Awaya[2] · Daisuke Kurisu[3]

Abstract

We investigate the estimation methods of the multivariate non-stationary errors-in-variables models when there are non-stationary trend components and the measurement errors or noise components. We compare the maximum likelihood (ML) estimation and the separating information maximum likelihood (SIML) estimation. The latter was proposed by Kunitomo and Sato (Trend, seasonality and economic time series: the nonstationary errors-in-variables models. MIMS-RBP-SDS-3, MIMS, Meiji University. http://www.mims.meiji.ac.jp/, 2017) and Kunitomo et al. (Separating information maximum likelihood method for high-frequency financial data. Springer, Berlin, 2018). We have found that the Gaussian likelihood function can have non-concave shape in some cases and the ML method does work only when the Gaussianity of non-stationary and stationary components holds with some restrictions such as the signal–noise variance ratio in the parameter space. The SIML estimation has the asymptotic robust properties in more general situations. We explore the finite sample and asymptotic properties of the ML and SIML methods for the non-stationary errors-in variables models.

✉ Naoto Kunitomo
   naoto.kunitomo@gmail.com

[1] School of Political Science and Economics, Meiji University, Sarugakucho 3rd Building C-106, 1-1 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-8301, Japan

[2] Department of Statistical Science, Duke University, P.O. Box 90251, Durham, NC 27708-0251, USA

[3] School of Engineering, Tokyo Institute of Technology, Meguro-ku, Ôokayama 2-12-1 W9-74, Tokyo 152-8552, Japan

1 Introduction

There has been a vast amount of published research on the use of statistical time series analysis of macro-economic time series. One important feature of macroeconomic time series, which may be different from the standard time series analysis, is the fact that the observed time series is an apparent mixture of non-stationary components and stationary components. The second feature is that the measurement errors in economic time series often play important roles because macro-economic data are usually constructed from various sources including sample surveys in major official statistics although their statistical analysis often ignores measurement errors. There is the third important issue that the sample size of macro-economic data is rather small and we have 120, say, time series observations for each series when we have quarterly data over 30 years. The quarterly GDP series, for instance, are published from 1994 data to present by the cabinet office of Japan. Since the sample size is small, it is important to use an appropriate statistical procedure to extract information on trend and noise (or measurement error) components in a systematic way from data. Some of these aspects have been discussed by Morgenstern (1950), Granger and Newbold (1977), and Nerlove et al. (1995) for instance. See https://www.esri.cao.go.jp/index-e.html for the official macro-economic (GDP) data published by Cabinet Office of Japanese Government.

Although economists often take the transformation of the non-stationary time series such as log-transformation and/or differencing by the Box–Jenkins method (Box and Jenkins 1970), the standard assumption on statistical procedures after joint transformation of multiple time series data may not be valid when we have measurement errors. In this regard, Kunitomo and Sato (2017, 2019) [or Kunitomo et al. (2018)] developed a new 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 analysis are Engle and Granger (1987) and Johansen (1995), which dealt with multivariate non-stationary and stationary time series and developed the notion of co-integration without measurement errors. The problem of our interest is related to their work, but it has different aspects and our focus is on the non-stationarity and measurement error in the non-stationary errors-in-variables models. Also in econometric literature the issue of identification of parametric models and the issue of estimation when the true parameters are around the boundary points have been discussed by Rothenberg (1971) and Andrews (1999). These works are potentially related to our problem in the non-stationary errors-in-variable models, which would be an interesting future topic.

In statistics literature, on the other hand, the state space modeling and filtering for non-stationary time series has been developed by Akaike (1989) and Kitagawa (2010) among other similar attempts and there have been applications in many fields in control engineering and statistical seismology already reported. Their method may look different from the framework of standard time series econometrics at the first glance, but the underlying statistical problem is essentially

the same because of the non-stationarity of time series and measurement errors. Our study can be regarded as an investigation of statistical inference problem of state space modeling, which is related to an earlier work by Chang et al. (2008), for instance. The problem of the ML estimation in our setting may be related to Anderson and Takemura (1986) in a different context.

In the statistical multivariate analysis, there is also some literature on the errors-in-variables models as Anderson (1984, 2003) and Fuller (1987), but they considered the multivariate cases of independent observations and the underlying situation is different from ours.

The main purpose of this paper is to compare the SIML estimation and the maximum likelihood (ML) estimation, which are two different methods to estimate the multivariate non-stationary errors-in-variables models when there are non-stationary trends and noise components. We investigate the finite and large sample properties of two estimation methods. An important finding is the fact that the Gaussian likelihood function may have non-concave shape in some case of the non-stationary errors-in-variables models although the ML method works well when the Gaussianities of non-stationary and stationary components hold with some restriction on the parameter space and under the assumption that the measurement errors are not small. When the measurement error is small in a sense and/or there are co-integrated relations among trends with the rank being smaller than the dimension of observations, there could be a serious problem in the ML estimation under the assumption of Gaussian distributions (see Sect. 3.1 on the details of this issue). The SIML method, on the other hand, gives an alternative way to overcome the underlying difficulty in a non-parametric way. It has the asymptotic robust properties under general conditions of moments for consistency and for asymptotic normality.

In Sect. 2 we present a general formulation of the non-stationary errors-in-variables models and explain two estimation methods, that is the SIML and ML methods in non-stationary time series. We also give simple examples to illustrate the importance of measurement error problem in non-stationary time series models and explain some motivations of the present study. In Sect. 3, as a simple example we use one-dimensional case with the random walk plus noise model and then we discuss a common (non-stationary) factor case as the two-dimensions errors-in-variables model. In Sect. 4 we investigate the Gaussian likelihood function and its shape. We give the consistency result of the ML estimation under Gaussianity, which may be new although it is not surprising. We also discuss the asymptotic properties and simulation results on the SIML method. Then in Sect. 5, we shall discuss possible extensions of our analysis and give concluding remarks in Sect. 6. Some mathematical derivations will be given in the Appendix.

2 Non-stationary errors-in-variables models

2.1 The basic formulation and estimation methods

Let y_{ji} be the i-th observation of the j-th time series at i for $i = 1, \ldots, n; j = 1, \ldots, p$. We set $\mathbf{y}_i = (y_{1i}, \ldots, y_{pi})'$ be a $p \times 1$ vector and $\mathbf{Y}_n = (\mathbf{y}_i')\ (= (y_{ij}))$ be an $n \times p$ matrix

of observations and denote $\mathbf{y}_0$ as the initial $p \times 1$ vector. We consider the situation when the underlying non-stationary trends $\mathbf{x}_i \, (= (x_{ji})) \, (i = 1, \ldots, n)$ are not necessarily the same as the observed time series and let $\mathbf{v}_i' = (v_{1i}, \ldots, v_{pi})$ be the vector of noise components, which are independent of $\mathbf{x}_i$. Then we use the additive state space decomposition form as [see Akaike (1989) and Kitagawa (2010)]

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

where $\mathbf{x}_i \, (i = 1, \ldots, n)$ are a sequence of non-stationary trend components satisfying

$$\Delta \mathbf{x}_i = (1 - \mathcal{L})\mathbf{x}_i = \mathbf{w}_i^{(x)} \tag{2}$$

with $\mathcal{L}\mathbf{x}_i = \mathbf{x}_{i-1}$, $\Delta = 1 - \mathcal{L}$, $\mathcal{E}(\mathbf{w}_i^{(x)}) = \mathbf{0}$, $\mathcal{E}(\mathbf{w}_i^{(x)}\mathbf{w}_i^{(x)'}) = \Sigma_x$, and $\mathbf{v}_i \, (i = 1, \cdots, n)$ are a sequence of (mutually) independent noise components with $\mathcal{E}(\mathbf{v}_i) = \mathbf{0}$, $\mathcal{E}(\mathbf{v}_i\mathbf{v}_i') = \Sigma_v$.

We assume that $\mathbf{w}_i^{(x)}$ and $\mathbf{v}_i$ are the sequence of i.i.d. random variables with Σ_v being non-negative definite and finite, and the random variables $\mathbf{w}_i^{(x)}$ and $\mathbf{v}_i$ are mutually independent.

We consider the situation when $\Delta \mathbf{x}_i$ and $\mathbf{v}_i \, (i = 1, \ldots, n)$ are mutually independent and each of the component vectors are independently, identically, and normally distributed as $N_p(\mathbf{0}, \Sigma_x)$ and $N_p(\mathbf{0}, \Sigma_v)$, respectively. We use an $n \times p$ matrix $\mathbf{Y}_n = (\mathbf{y}_i')$ and consider the distribution of $np \times 1$ random vector $(\mathbf{y}_1', \ldots, \mathbf{y}_n')'$. Given the initial condition $\mathbf{y}_0$, we have

$$\mathrm{vec}(\mathbf{Y}_n) \sim N_{n \times p}(\mathbf{1}_n \cdot \mathbf{y}_0', \mathbf{I}_n \otimes \Sigma_v + \mathbf{C}_n\mathbf{C}_n' \otimes \Sigma_x), \tag{3}$$

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

$$\mathbf{C}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 0 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}_{n \times n}. \tag{4}$$

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

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

where

$$\bar{\mathbf{Y}}_0 = \mathbf{1}_n \cdot \mathbf{y}_0'. \tag{5}$$

To develop the method of the SIML estimation [see Kunitomo and Sato (2017)], we use the K_n-transformation from $\mathbf{Y}_n$ to $\mathbf{Z}_n \, (= (\mathbf{z}_k'))$ by

$$\mathbf{Z}_n = \mathbf{K}_n(\mathbf{Y}_n - \bar{\mathbf{Y}}_0), \quad \mathbf{K}_n = \mathbf{P}_n\mathbf{C}_n^{-1}, \tag{6}$$

where

$$\mathbf{C}_n^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}_{n \times n}, \tag{7}$$

and

$$\mathbf{P}_n = (p_{jk}^{(n)}), \; p_{jk}^{(n)} = \sqrt{\frac{2}{n + \frac{1}{2}}} \cos \left[\frac{2\pi}{2n+1} \left(k - \frac{1}{2} \right) \left(j - \frac{1}{2} \right) \right]. \tag{8}$$

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 $d_k = 2[1 - \cos(\pi(\frac{2k-1}{2n+1}))]$ $(k = 1, \dots, n)$. Then the conditional likelihood function given the initial condition is proportional to

$$L_n = \sum_{k=1}^n \log |a_{kn}^* \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_x|^{-1/2} - \frac{1}{2} \sum_{k=1}^n \mathbf{z}_k' [a_{kn}^* \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_x]^{-1} \mathbf{z}_k, \tag{9}$$

where

$$a_{kn}^* \; (= d_k) = 4 \sin^2 \left[\frac{\pi}{2} \left(\frac{2k-1}{2n+1} \right) \right] \quad (k = 1, \dots, n). \tag{10}$$

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

The ML estimation of unknown parameters is defined as the maximization of (9) with respect to $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_x$. Since the coefficients a_{kn}^* $(k = 1, \dots, n)$, the ML estimator is a complicated function of data and its computation is not a trivial task as we shall see in Sects. 3 and 4.

From the representation (9), it may be natural to use $\mathbf{z}_k \mathbf{z}_k'$ to estimate $a_{kn}^* \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_x$ 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 $\hat{\boldsymbol{\Sigma}}_x$ is defined by

$$\hat{\boldsymbol{\Sigma}}_{x,SIML} = \frac{1}{m_n} \sum_{k=1}^{m_n} \mathbf{z}_k \mathbf{z}_k'. \tag{11}$$

---

[1] We have used a_{kn}^*, which is slightly different a_k in Kunitomo et al. (2018) and the latter corresponds to $a_{kn} = n a_{kn}^*$. They have investigated the SIML estimation when the length of observation interval decreases in a fixed period as the high-frequency asymptotics.

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 using only the trend parts without measurement errors from the time series observations. The interpretation of (11) from the frequency domain of non-stationary time series will be discussed in a general setting [see (42) and (43) in Sect. 5].

From our construction of the SIML estimation the essential features of estimation do not much depend on the presence of noise terms when the noise terms are stationary. This feature was the main reason for developing the SIML method by Kunitomo and Sato (2017) and Kunitomo et al. (2018).

Let the quadratic variation of observed vectors $\mathbf{y}_i$ $(i = 1, \dots, n)$ be

$$\mathbf{QV}_y(1) = \sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{y}_{i-1})(\mathbf{y}_i - \mathbf{y}_{i-1})', \tag{12}$$

where $\mathbf{y}_0$ is the initial vector. If we denote $\hat{\boldsymbol{\Sigma}}_y = (1/n)\mathbf{QV}_y(1)$, we need the condition $\hat{\boldsymbol{\Sigma}}_{x,\text{SIML}} \leq \hat{\boldsymbol{\Sigma}}_y$ in the sense of positive semi-definiteness.

For the SIML estimation $\hat{\boldsymbol{\Sigma}}_{x,\text{SIML}}$, the number of terms m_n should be dependent on n. In the representation of (11) 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. There is a trade-off between the bias and the variances of the SIML estimator for $\boldsymbol{\Sigma}_x$. Kunitomo and Sato (2017) have shown that when $m_n \to \infty$ as $n \to +\infty$ for the consistency we need the condition $0 < \alpha < 1.0$ while for the asymptotic normality we need the condition $0 < \alpha < 0.8$ under the assumption that the parameter matrices $\boldsymbol{\Sigma}_x > 0$ and $\boldsymbol{\Sigma}_v > 0$ are fixed (see Theorem 4.3 in Sect. 4). As an example, we may take $\alpha = 0.79$.

2.2 Non-stationary time series and measurement errors

In this subsection using a simple example when $p = 1$ we shall illustrate the reason why the presence of noise term in the non-stationary time series, even if it is small, forces us to change the standard thinking on time series analysis.

In traditional econometric analysis of time series, the non-stationarity of economic times series has been often discussed, but there were not many discussions on the role of measurement errors [see Engle and Granger (1987) and Johansen (1995) for instance]. The standard arguments for the integrated processes have been to use the Brownian functionals to describe the behaviors of integrated processes when the sample size is large. As a typical example,[2] we generate a set of mutually independent variables $w_i^{(x)}$ $(i = 1, \dots, n)$, which follows $\sqrt{12}(U(0, 1) - 0.5)$ and $U(0, 1)$ follows the uniform distribution. Then we generate x_i and y_i satisfying $x_i = x_{i-1} + w_i^{(x)}$, and $y_i = x_i + v_i$ $(i = 1, \dots, n)$ by adding the measurement errors v_i, which are independent $N(0, 1)$ random variables. By replicating 1000 times, the empirical

---

[2] We have followed a suggestion from a referee of JJSD on this example, which corresponds to the simple case in Sect. 3.1.
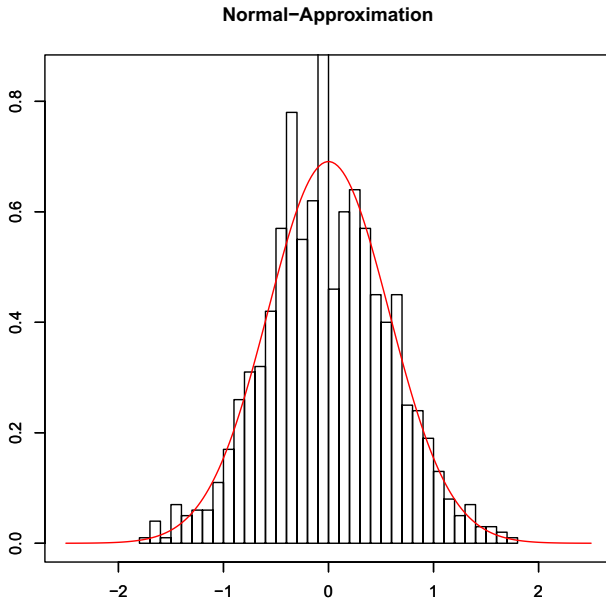
Normal−Approximation

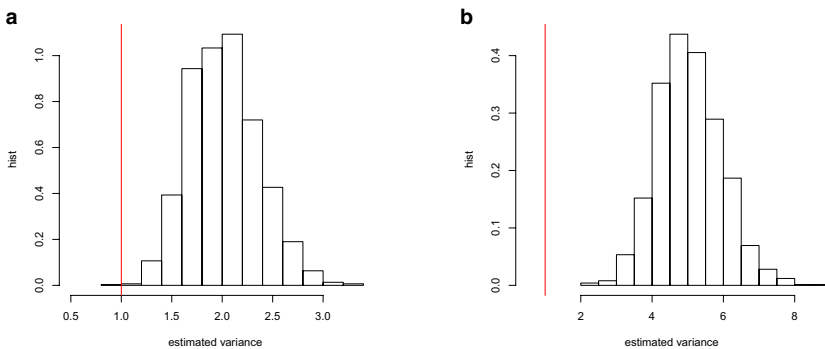**Fig. 1** Normal approximation ($n = 30$)

Fig. 2 (**a**) (Left) Estimated variance ($n = 80, c = 2$), (**b**) (right) estimated variance ($n = 80, c = 1/2$)

distribution of $[\sum_{i=1}^{n} y_i]/[\sigma_x \sqrt{n^3}]$ (we have set $\sigma_x^2 = \mathcal{E}(w_i^{(x)2}) = 1$ and $x_0 = 0$) can be approximated by the weak convergence to the Brownian functional

$$X(1) = \int_0^1 B(s)\,\mathrm{d}s, \tag{13}$$

where $B(s)$ is the Brownian motion on [0, 1] and it follows $N(0, 1/3)$. Figure 1 illustrates the fact that this approximation is reasonable even when $n = 30$. (The red curve
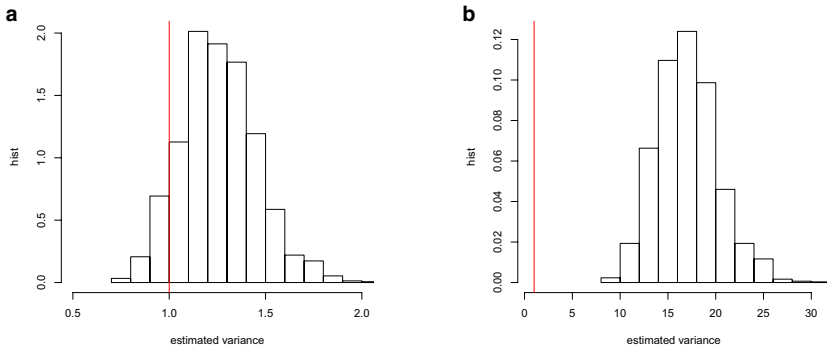
Fig. 3 (**a**) (Left) Estimated variance ($n = 80, c = 8$), (**b**) (right) estimated variance ($n = 80, c = 1/8$)

shows $N(0, 1 / 3)$.) However, as illustrated in Chapters 9 and 10 of Hayashi (2000), to make statistical inference on the unit roots and possible co-integrated vectors for instance, we need an estimate of the variance and covariances of innovation terms of integrated processes such as σ_x^2, which are generally unknown. There should be a careful analysis on the role of innovations, but the easiest way to estimate the variance of innovation term of the integrated processes and to use the normalized sum of squared differences of observed time series as $\hat{\sigma}_y^2 \ (= (1/n) \sum_{i=2}^{n}(y_i - y_{i-1})^2)$ although there can be more sophisticated methods such as the maximum likelihood method in more general situations. As a simple illustration, we generate a set of observations $y_i \ (i = 1, \ldots, n)$ by adding the measurement errors v_i which are independent $N(0, 0.5)$ random variables to the I(1) process x_i, and thus $y_i = x_i + v_i \ (i = 1, \ldots, n)$. Figure 2a (case 2-1) illustrates the estimated variance of innovation $w_i^{(x)} \ (= x_i - x_{i-1})$ by this method, where the true parameter value (as red line) is 1. Although the variance of innovation of the integrated process is twice of the variance of measurement errors and the measurement errors is small in a sense, the standard estimated value has a significant bias and it is distributed around 2 when $n = 80$. This case corresponds to $c = 2 = \sigma_x^2/\sigma_v^2$. We also give Figs. 2b, 3a, b for case 2-2 ($c = 1/2$), case 2-3 ($c = 8$) and case 2-4 ($c = 1/8$) for a comparison. Case 2-1 and case 2-3 correspond to the small noise cases while case 2-2 and case 2-4 correspond to the large noise cases. In latter cases the bias of the estimated variance σ_x^2 becomes large as we can expect.

The point here is the fact that even when we have small (Gaussian) noise we may have misleading estimation on the variance of system variables. These examples may illustrate the importance of our analysis of measurement errors in non-stationary time series. We usually do not have much information on the magnitude of c, which is the ratio of the signal variance over the noise variance, and the distribution of measurement errors in advance when we observe non-stationary data. Hence it is important to use the statistical method which does not depend on the value of c and the distribution of noise for practical purpose.

3 Simple cases

3.1 An illustrative example

To see the main problem of our interest in a clear way, we consider the simplest case when $p = 1$. Let y_i be the ith observation of time series for $i = 1, \ldots, n$ 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 ($i = 1, \ldots, n$) satisfies

$$x_i = x_{i-1} + w_i^{(x)}, \tag{14}$$

where $w_i^{(x)}$ are the independent random variables followed by $N(0, \sigma_x^2)$ and x_0 is the initial variable. Let v_i (measurement error) be the sequence of i.i.d. random variables followed by $N(0, \sigma_v^2)$, which are independent of x_i ($i = 1, \ldots, n$). For the additive model

$$y_i = x_i + v_i \quad (i = 1, \ldots, n), \tag{15}$$

the log-likelihood function is proportional to

$$L_n = \sum_{k=1}^{n} \log |a_{kn}^* \sigma_v^2 + \sigma_x^2|^{-1/2} - \frac{1}{2} \sum_{k=1}^{n} \frac{z_k^2}{a_{kn}^* \sigma_v^2 + \sigma_x^2}, \tag{16}$$

where

$$a_{kn}^* = 4 \sin^2 \left[\frac{\pi}{2} \left(\frac{2k-1}{2n+1} \right) \right] \quad (k = 1, \ldots, n). \tag{17}$$

Using the variance ratio $c = \sigma_x^2 / \sigma_v^2 \ (\geq 0)$,[3] we rewrite $-(1/2)L_n$ as

$$L_{1n} = \sum_{k=1}^{n} [\log \sigma_v^2 + \log(a_{kn}^* + c)] + \frac{1}{\sigma_v^2} \sum_{k=1}^{n} \frac{z_k^2}{a_{kn}^* + c}. \tag{18}$$

Since $z_k \sim N(0, a_{kn}^* \sigma_v^2 + \sigma_x^2)$ ($k = 1, \ldots, n$), the maximum likelihood estimator of σ_v^2 can be represented as

$$\hat{\sigma}_{v.ML}^2 = \frac{1}{n} \sum_{k=1}^{n} \frac{z_k^2}{a_{kn}^* + c} \tag{19}$$

and the concentrated (normalized) log-likelihood function in this simple case is proportional to $-(1/2)$ times

$$L_{1n}(c) = \log \left[\frac{1}{n} \sum_{k=1}^{n} \frac{z_k^2}{a_{kn}^* + c} \right] + \frac{1}{n} \sum_{k=1}^{n} \log[a_{kn}^* + c] + 1. \tag{20}$$

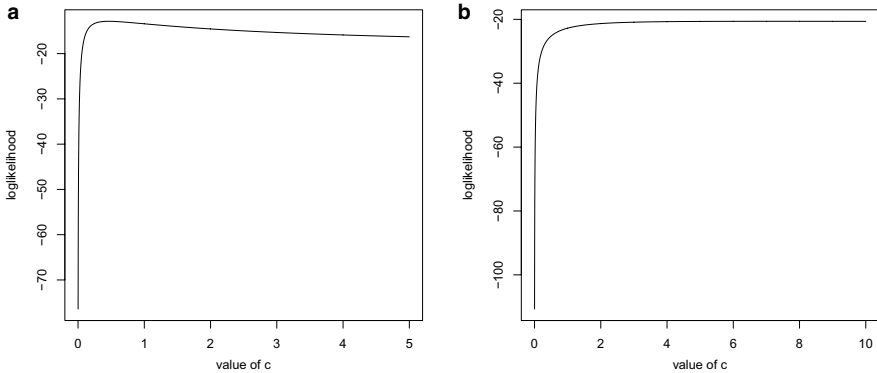[3] The present notation c corresponds to $c^{-1/2}$ in Akaike (1989).

Fig. 4 **(a)** (Left) Likelihood function ($n = 100, c = 1/2$), **(b)** (right) likelihood function ($n = 100, c = 8$)
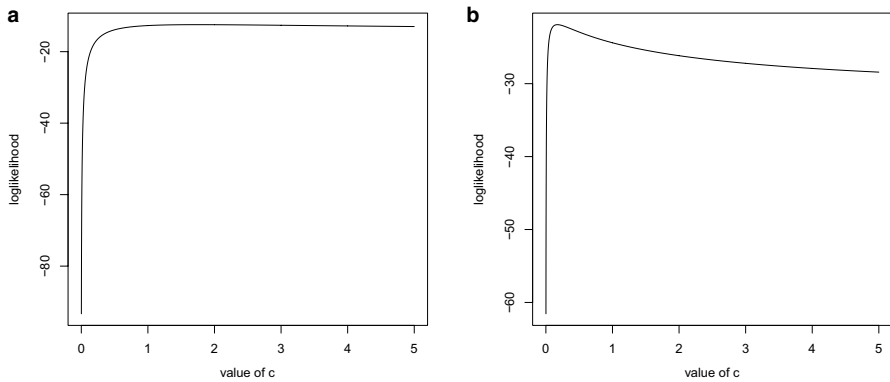

Fig. 5 **(a)** (Left) Likelihood function ($n = 100, c = 2$), **(b)** (right) likelihood function ($n = 100, c = 1/8$)

Then it may not be straight-forward to obtain the maximum likelihood estimator of c because the likelihood function may not be a simple function and the likelihood equation $\frac{\partial L_{1n}(c)}{\partial c} = 0$ is a polynomial function of order $2n - 3$, and as consequence there are local maximum points with any finite sample data.

As a typical situation we draw the likelihood function with respect to the parameter c and the result of a small simulation in Fig. 4a when the true values are $\sigma_v^2 = 0.4$ and $\sigma_x^2 = 0.2$ (case 3-1). We also show the likelihood function and the result of case when the true values are $\sigma_v^2 = 0.1$ and $\sigma_x^2 = 0.8$ (case 3-2) as Fig. 4b. To make a comparison, we give case 3-3 ($\sigma_v^2 = 0.2, \sigma_x^2 = 0.4$) as Fig. 5a and case 3-4 ($\sigma_v^2 = 0.8, \sigma_x^2 = 0.1$ as Fig. 5b. The shape of the likelihood function is reasonable in case 3.1 and case 3.4 while it is not so in case 3.2 and case 3.3. In the latter cases, the likelihood function is rather flat around the maximum point, which means that it does not have much information in a sense.

Case 3-1 and case 3-4 are the standard ones and we can expect that the ML estimation under the Gaussian assumption gives a reasonable result. However, case 3-2
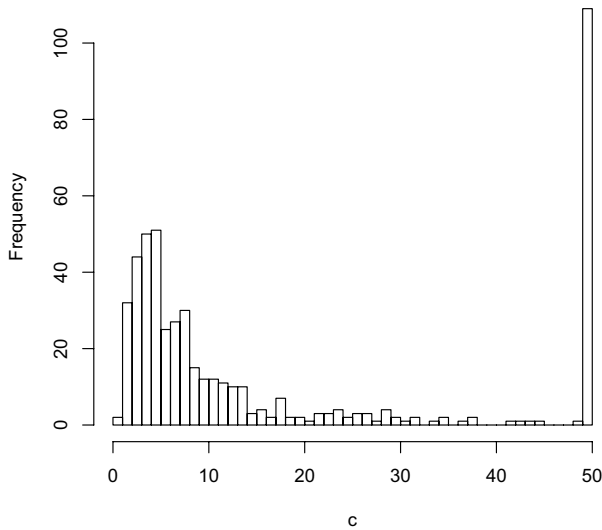
Fig. 6 Histogram of ML ($n = 100, c = 8$)

and case 3.3 illustrate some problem on the ML estimation. The likelihood function looks flat over a wide range of the parameter space of c and it causes some computation difficulty to find its maximum. We could interpret the reason why Akaike (1989) suggested that we should impose the restriction $0 < c \leq c_u$ for a pre-specified c_u in our setting because of the usefulness of the resulting statistical models of time series filtering. Kitagawa (2010) has developed DECOMP program,[4] which has a similar restriction on the parameter space. We sometimes obtain the estimated value quite near 1.0 for c when we analyze macro-economic data such as real (quarterly) GDP in Japan.

One of interesting aspects of the present problem is the fact the ML method does not necessarily give a satisfactory solution meaning a numerically stable solution and it is the case when $p = 1, \sigma_v^2 = 0.1$ and $\sigma_x^2 = 0.8$. Since the likelihood function is near flat over a wide range of the parameter space, it is often difficult to find the maximum in a stable way. In fact, given the finite sample, there is a positive probability of zero when the true value of σ_v^2 is small, but not zero. This can be easily shown because the normalized second-order derivative of the log-likelihood function can be approximated as some negative number when $\tau (= 1/c) \to 0$ and the zero point is a local maximum. As an illustration, we give the histogram of the ML estimation with the restriction $0 < c \leq 50$ for the second case as Fig. 6 with 500 replications. There is a difficulty to find the global maximum of likelihood function and there is some probability at the boundary point of $c = 50$.

---

[4] It is available freely at http://ssnt.ism.ac.jp/inets/inets_eng.html (Institute of Statistical Mathematics, Tokyo).

On the other hand, it is certainly possible to approximate the log-likelihood function (16) as

$$L_n^{SI} = \sum_{k=1}^{m_n} \log |a_{kn}^* \sigma_v^2 + \sigma_x^2|^{-1/2} - \frac{1}{2} \sum_{k=1}^{m_n} \frac{z_k^2}{a_{kn}^* \sigma_v^2 + \sigma_x^2}. \tag{21}$$

We set the requirement on $m_n/n = o(1)$ because $a_{kn}^* = o(1)$ when $k = 1, \ldots, m_n$ and $m_n, n \to \infty$. Then the SIML estimator σ_x^2 (we call the macro-SIML method) can be given by

$$\hat{\sigma}_{x.\text{SIML}}^2 = \frac{1}{m_n} \sum_{k=1}^{m_n} z_k^2, \tag{22}$$

Since the information on the trend term is separated from the noise term, we expect that the resulting macro-SIML estimation has a robust property.

We note that the macro-SIML estimation of σ_v^2 is not the same as the original (finance) SIML method developed by Kunitomo et al. (2018) because $a_{kn}^* = O(1)$ for $k = n - m_n + 1, \ldots, n$. One way to estimate σ_v^2 is to use the fact that

$$\mathcal{E}\left[\frac{1}{n} \sum_{k=1}^n z_k^2\right] = \sigma_x^2 + \left(\frac{1}{n} \sum_{k=1}^n a_{kn}^*\right)\sigma_v^2 \longrightarrow \sigma_x^2 + 2\sigma_v^2 \quad (n \to \infty). \tag{23}$$

It is clear that $a_{kn}^* \longrightarrow 0$ 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

$$\hat{\sigma}_{v.\text{SIML}}^2(1) = \frac{1}{2}\left[\frac{1}{n} \sum_{k=1}^n z_k^2 - \hat{\sigma}_{x.\text{SIML}}^2\right], \tag{24}$$

with the restriction of non-negativity.

For the estimation problem of high-frequency financial data, Kunitomo et al. (2018) have suggested to use $\hat{\sigma}_{v.\text{SIML}}^2 = \frac{1}{l_n} \sum_{k=n-l_n}^n a_{kn}^{-1} z_k^2$, where $a_{kn} = n a_{kn}^*$ and $l_n = o(n)$ for the high-frequency asymptotics. However, in the present case of Macro-SIML with a_{kn}^* it is straight-forward to show that $a_{n-l+1,n}^* \to 4$ as $n \to \infty$ for fixed l, the macro-SIML estimator is given by

$$\hat{\sigma}_{v.\text{SIML}}^2(2) = \frac{1}{l_n} \sum_{k=n-l_n+1}^n a_{kn}^{*-1} z_k^2 - \frac{1}{4}\hat{\sigma}_{x.\text{SIML}}^2 \tag{25}$$

with the restriction of non-negativity.

3.2 A non-stationary common trend case

The difficulty in the ML estimation becomes more serious in the multivariate non-stationary errors-in-variables models. We illustrate the problem of multivariate

aspects in a simple, but important formulation. It can be regarded as a simple extension of the so-called reduced rank regression.

Let $\mathbf{y}_i$ be the ith observation of p-dimensional time series $(i = 1, \ldots, n)$, $\mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i$ and $\mathbf{Y}_n = (\mathbf{y}_i')$ be an $n \times p$ $(p > 1)$ matrix of observation. We assume that the vectors $\mathbf{x}_i$ satisfy

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{w}_i^{(x)}, \tag{26}$$

and $\mathbf{w}_i^{(x)} = \boldsymbol{\pi}\mu_i^*$, $\boldsymbol{\pi}$ is a (non-zero) $p \times 1$ vector, μ_i^* is a sequence of i.i.d. (one-dimensional) random variables[5] 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 (non-singular) matrix $\boldsymbol{\Sigma}_v$. We set $\mathbf{b} = \sigma_\mu \boldsymbol{\pi}$ and $\mathbf{A} = a_{kn}^* \boldsymbol{\Sigma}_v$, and then apply the matrix formulas such that for a positive definite $\mathbf{A}$ and non-zero vector $\mathbf{b}$

$$|\mathbf{A} + \mathbf{bb}'| = |\mathbf{A}|[1 + \mathbf{b}'\mathbf{A}^{-1}\mathbf{b}] \tag{27}$$

and

$$[\mathbf{A} + \mathbf{bb}']^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{b}[1 + \mathbf{b}'\mathbf{A}^{-1}\mathbf{b}]^{-1}\mathbf{b}'\mathbf{A}^{-1} \tag{28}$$

for $\boldsymbol{\Sigma}_x = \mathbf{bb}'$.

The likelihood function L_n is proportional to $(-1/2)$ times

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

where we denote

$$c = \sigma_\mu^2 \boldsymbol{\pi}' \boldsymbol{\Sigma}_v^{-1} \boldsymbol{\pi}. \tag{29}$$

We need a normalization for vector $\boldsymbol{\pi}$ and one possibility is to take $\boldsymbol{\pi}' = (1, -\boldsymbol{\theta}_2')$, but there can be other possibility.

---

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

Remark 3.1 When $p = 2$, we take $\boldsymbol{\beta}' = (1, -\beta_2)$. and $\boldsymbol{\pi}'\boldsymbol{\beta} = 0$ with a normalization. Then we can interpret $\boldsymbol{\beta}'\mathbf{y}_i = \boldsymbol{\beta}'\mathbf{v}_i (= u_i)$ (the rank of $\boldsymbol{\pi}$ is 1) as the structural equation in time series econometrics. It is because $\mathbf{y}_i$ is an I(1) vector and while $\boldsymbol{\beta}'\mathbf{y}_i = u_i$ is an I(0) variable, where d in I(d) ($d = 0, 1$) is the integration order of time series.

As an intuitive way to simplify the present problem of statistical relationship among non-stationary variables and to obtain the solution is to use the moment condition that

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

and

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

In the present case, the rank of matrix $\boldsymbol{\Sigma}_x$ is one while the matrix $\boldsymbol{\Sigma}_v$ has a full rank. When $p = 2$ in particular, we can find a vector $\boldsymbol{\beta}$ uniquely such that $\boldsymbol{\Sigma}_x\boldsymbol{\beta} = 0$ with a normalization (see Sect. 5.2 for more general cases).

To estimate the structural equation vector $\boldsymbol{\beta}$, then, it may be natural to consider the characteristic equation

$$[\hat{\boldsymbol{\Sigma}}_{x.\text{SIML}} - \lambda\hat{\boldsymbol{\Sigma}}_{v.\text{SIML}}]\hat{\boldsymbol{\beta}}_{\text{SIML}} = \mathbf{0}, \tag{30}$$

and

$$\hat{\boldsymbol{\Sigma}}_{x.\text{SIML}} = \frac{1}{m_n} \sum_{k=1}^{m_n} \mathbf{z}_k\mathbf{z}_k', \tag{31}$$

$$\hat{\boldsymbol{\Sigma}}_{v.\text{SIML}}(1) = \frac{1}{2}\left[\frac{1}{n}\sum_{k=1}^{n}\mathbf{z}_k\mathbf{z}_k' - \hat{\boldsymbol{\Sigma}}_{x.\text{SIML}}\right], \tag{32}$$

or

$$\hat{\boldsymbol{\Sigma}}_{v.\text{SIML}}(2) = \frac{1}{l_n}\sum_{k=n+1-l_n}^{n} a_{kn}^{*-1}\mathbf{z}_k\mathbf{z}_k' - \frac{1}{4}\hat{\boldsymbol{\Sigma}}_{x.\text{SIML}}, \tag{33}$$

where $\mathbf{Z}_n = (\mathbf{z}_k') = \mathbf{P}_n\mathbf{C}_n^{-1}(\mathbf{Y}_n - \mathbf{1}_n\bar{\mathbf{y}}_0')$ as (6), λ is the (scalar) eigenvalue, and $\hat{\boldsymbol{\Sigma}}_{x.\text{SIML}}$ and $\hat{\boldsymbol{\Sigma}}_{v.\text{SIML}}$ are the SIML estimators of $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_v$, respectively. We require the condition that these estimators of variance–covariance matrices are non-negative definite.

Since the rank of $\boldsymbol{\Sigma}_x$ is degenerated and it is one in the present case, it may be natural to use the smaller eigenvalue, say, λ_1. Then the resulting characteristic vector $\hat{\boldsymbol{\beta}}_{\text{SIML}}$, is called the SIML estimator of $\boldsymbol{\beta}$ because of (30). Since the estimated variance–covariance matrix of $\boldsymbol{\Sigma}_v$ should be positive definite, we may have instability in some cases if we use (32) or (33) without any restriction as the non-negative definiteness.

Fig. 7 Likelihood function of θ ($n = 1000$)
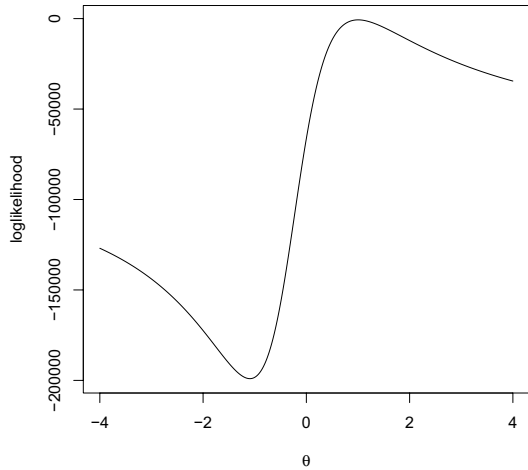
A simplified (consistent) estimation may be given by

$$\hat{\boldsymbol{\Sigma}}_{x.\text{SIML}} \times \hat{\boldsymbol{\beta}}_{\text{SIL}} = \mathbf{0}, \tag{34}$$

that is,

$$\hat{\boldsymbol{\Sigma}}_{x.\text{SIML}} \times \begin{bmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.\text{SIL}} \end{bmatrix} = \mathbf{0}. \tag{35}$$

We can solve as

$$\hat{\boldsymbol{\beta}}_{2.\text{SIL}} = \hat{\boldsymbol{\Sigma}}_{22x.\text{SIML}}^{-1} \hat{\boldsymbol{\Sigma}}_{21x.\text{SIML}}, \tag{36}$$

where $\hat{\boldsymbol{\Sigma}}_{22x.\text{SIML}}$ and $\hat{\boldsymbol{\Sigma}}_{21x.\text{SIML}}$ are the (2, 2)-element and (2, 1)-element of $\hat{\boldsymbol{\Sigma}}_{x.\text{SIML}}$, respectively. ($\hat{\boldsymbol{\Sigma}}_{22x.\text{SIML}}$ is positive with probability one.)

It is the least squares method for the transformed variables $\mathbf{z}_k$ ($k = 1, \ldots, m$) and hence we call the separating information least squares (SILS) estimator.

Remark 3.2 We note that there were extensive discussions on alternative estimation methods including the similar form as the SIML and the SISL estimators for the errors-in-variables models and the single structural equation econometric models for independent observations. Some improvements on the finite sample properties may be possible. See Anderson (1984), Anderson et al. (1986), and Fuller (1987).

4 Gaussian likelihood function and estimation methods

It may be natural to apply the general parametric principle of the maximum likelihood (ML) method. One of interesting aspects of the present problem is the fact the ML method does not necessarily give a satisfactory solution.

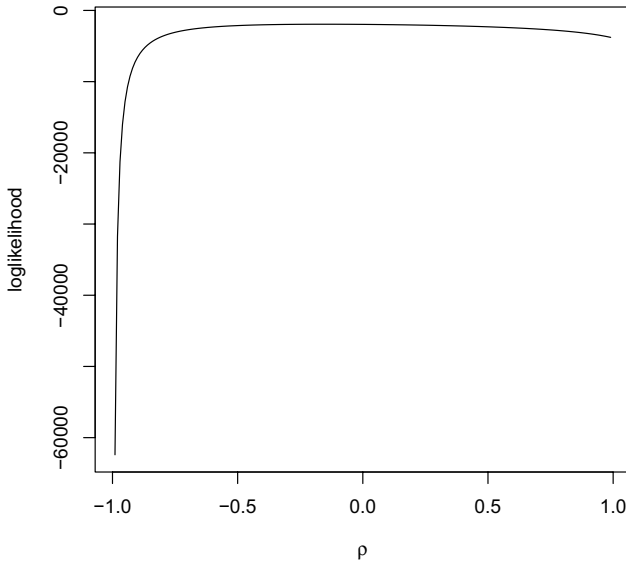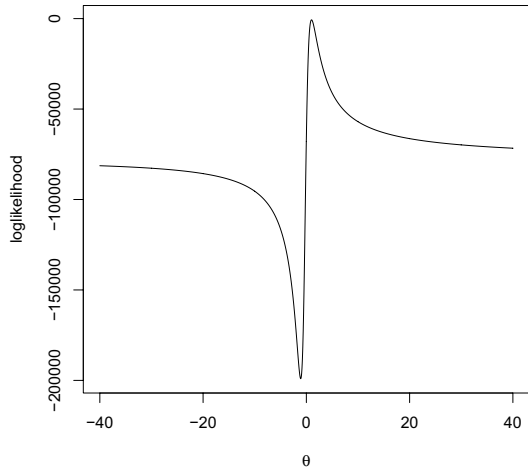**Fig. 8** Likelihood function of θ
($n = 1000$)

Fig. 9 Wrong likelihood function of ρ ($n = 1000$)

As a two-dimensions example, we use the example in Sect. 3.2 and set the true parameter values as $\sigma_\mu^2 = 0.4, \theta = 1.0$ and

$$\Sigma_v = \begin{pmatrix} 0.45 & 0.23 \\ 0.23 & 0.4 \end{pmatrix}, \quad \Sigma_x = \sigma_\mu^2 \pi \pi', \quad \pi = \begin{pmatrix} 1 \\ -\theta \end{pmatrix}.$$

Then we generate a set of simulated observations and we have drawn the Gaussian likelihood functions of θ in Figs. 7 and 8 when $n = 1000$, given the true values for other parameters. It is possible to attain the maximum point of the likelihood

function locally as shown in Fig. 7. Figure 8 is the same figure and suggests that there is a global maximization problem because we need the right starting point for the maximization. We have investigated the likelihood function in different cases and found that the likelihood function could have some non-concave forms in some cases as illustrated by Fig. 8. We found that the Gaussian likelihood function is nearly flat with respect to the correlation coefficient parameter and the maximization may be difficult with respect to correlation of the noise terms. These are some of important consequences in the non-stationary errors-in-variables models.

It is important to see what happens if the Gaussian assumption is not true and as an illustration we have drawn one wrong likelihood function in Fig. 9 on this problem. We generated the random variables followed by the uniform distribution on $[-2, +2]$ (i.e., $\mathbf{w}_i^{(x)} = (w_{ji}^{(x)})$), the distribution of $\mathbf{v}_i$ is normal and the correlation coefficient ρ is 0.3. As Fig. 9 (ρ, which is the correlation coefficient among measurement errors $\mathbf{v}_i$) suggests, the ML estimation of the variance-covariance matrix of trend components crucially depends on the assumption of Gaussianity as we had expected. Hence we have a risk to use the ML computation to investigate the relationships among hidden trend variables unless we have the strong support for the Gaussianity of data.

Now we investigate the asymptotic properties of $(-1)\times$ the log-likelihood function and the estimation methods (the ML and SIML estimators) when $\Sigma_x = \mathbf{bb}'$ ($\mathbf{b} \neq \mathbf{0}$) (i.e., the rank of Σ_x is 1) and $p \geq 2$. We normalize the Gaussian log-likelihood function by $-(1/n)$ and rewrite

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

$$(37)$$

where $\Sigma_v(\theta_0)$ and $\mathbf{b}(\theta_0)$ are Σ_v and $\mathbf{b}$, respectively, 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 $p \times p$ matrices $\mathbf{A}_k$,*

$$\mathcal{E}[(\mathrm{tr}(\mathbf{A}_k \mathbf{z}_k \mathbf{z}_k'))^2] = [\mathrm{tr}(\mathbf{A}_k \mathbf{Q})]^2 + 2\mathrm{tr}(\mathbf{A}_k \mathbf{Q} \mathbf{A}_k \mathbf{Q}). \tag{38}$$

Using this lemma, it is straight-forward to see that as $n \longrightarrow \infty$ the third term converges to

$$L_{12n}^* \xrightarrow{p} 0. \tag{39}$$

Then we can establish the next result on the ML estimation by evaluating the remaining terms of L_{1n}^* as in the 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 we have given it in the Appendix.

Theorem 4.2 *For $p \geq 2$ and the rank of Σ_x being 1, we set $\Sigma_x = \mathbf{bb}'$ in (1) and (2). Assume that $\mathbf{v}_i$ $(i = 1, \ldots, n)$ are the sequence of i.i.d. random vectors and Σ_v is a positive definite matrix. Then under the assumption of Gaussian distributions of $\mathbf{v}_i$ and $\mathbf{w}_i^{(x)}$, the maximum likelihood estimator of $\mathbf{b}$ and Σ_v are 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 using the standard regression method if the observed variables are co-integrated. Johansen (1995) had developed the ML method without any noise term and investigated the estimation and inference of the co-integrating vectors.

We should mention to the fact that the SIML estimator does have not only consistency as well as the asymptotic normality under the standard regularity conditions such as the 4-th order moments without the Gaussian assumption.

Theorem 4.3 (Kunitomo and Sato 2017) *For $p \geq 1$ Assume the non-stationary errors-in-variables model given by (1) and (2) with $\mathcal{E}[v_{ji}^4] < +\infty$ and $\mathcal{E}[w_{ji}^{(x)4}] < +\infty$ for $\mathbf{w}_i^{(x)} = (w_{ji}^{(x)})$ $(j = 1, \ldots, p; i = 1, \ldots, n)$. We set $m_n = [n^\alpha]$ $(0 < \alpha < 1)$. Then, the SIML estimator of Σ_x is consistent for $0 < \alpha < 1.0$ and asymptotically normal for $0 < \alpha < 0.8$ as $m_n \longrightarrow \infty$ $(n \longrightarrow \infty)$.*

We omit the proof because it is parallel to the one given in Chapter 5 of Kunitomo et al. (2018) and the detail is given in Kunitomo and Sato (2017).

There has not been any finite sample result on the estimation methods for the non-stationary time series models with errors-in-variables. Since there are many situations with macro-economic variables that we observe the non-stationarity and measurement errors, it is worthwhile to investigate the related issues using simulations. As we have seen in Sect. 3 the ML estimator has finite sample instability and we only report the finite sample properties of the SIML estimation in this section.

We have set $\sigma_\mu^2 = 1, \sigma_v^2 = 0.5, 2$ or 4, $\beta_2 = 1.5$, where we summarize our setting of simulations as

Table 1 Finite sample properties of SIML

σ_v^2	α	n	$\Sigma_{x,11}$	$\Sigma_{x,12}$	$\Sigma_{x,22}$	$\Sigma_{v,11}(1)$	$\Sigma_{v,12}(1)$	$\Sigma_{v,22}(1)$	$\beta_{2.\text{SIL}}$	$\beta_{2.\text{SIML}}$
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)

$$\Sigma_x = \begin{pmatrix} \Sigma_{x,11} & \Sigma_{x,12} \\ \Sigma_{x,12} & \Sigma_{x,22} \end{pmatrix} = \sigma_\mu^2 \begin{pmatrix} \beta_2 \\ 1 \end{pmatrix} (\beta_2, 1),$$

$$\Sigma_v = \begin{pmatrix} \Sigma_{v,11} & \Sigma_{v,12} \\ \Sigma_{v,12} & \Sigma_{v,22} \end{pmatrix}.$$

(The parametrization is slightly different from Figs. 7, 8, 9.) We took the cases when $n = 80$ or 400 and $m_n = [n^\alpha]$ with $\alpha = 0.6$ or 0.7, and the number of Monte Carlo repetition is 1500 in each case. From our simulations we summarize mail results as Table 1. In Table 1 the number inside the parentheses is the standard deviation of estimators calculated by our simulations. We found that the SIML estimator of trend variance–covariance estimates $\Sigma_{x.ij}$ $(i, j = 1, 2)$ have reasonable finite sample properties. We also give the SIML estimator of noise variance–covariance estimates $\Sigma_{v.ij}$ $(i, j = 1, 2)$, which is based on (22) are reasonable. The SILS estimate of β_2 is biased a little in comparison with the SIML estimates, but the former has smaller sample variance than the latter.

We have done many simulations, but the results are similar with Table 1 in the present formulation. There are several general findings, which are summarized as follows. First, on the effects of sample sizes the performance of the estimators of the SIML estimation, it becomes better as the sample size increases as we had expected. Second, when the variances of noises are small, both the SIL 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 bias. The variability of the SIML estimate in terms of simulation variance is slightly larger than that of the SIL estimate. Third, when the variances of noises are not small, the SILS estimator has a significant bias.

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.

5 Extensions

There can be several extensions of the problem we have been investigating. We discuss three of them rather briefly.

5.1 On autocorrelations of trend term

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 trend components with measurement errors. When there are autocorrelations in the trend terms in (2), the analysis of frequency domain on the underlying errors-in-variables models may give a new insight on the issue.

When the vector sequence of random vectors $\mathbf{w}_i^{(x)}$ $(i = 2, \dots, n)$ in (2) follows the stationary stochastic process, we write

$$\mathbf{w}_i^{(x)} = \sum_{j=0}^{\infty} \mathbf{C}_j^{(x)} \mathbf{e}_{i-j}^{(x)}, \tag{40}$$

where $\mathbf{C}_0^{(x)} = \mathbf{I}_p$, $\mathbf{C}_j^{(x)}$ $(j \geq 1)$ are absolutely convergent coefficient matrices and $\mathbf{e}_j^{(x)}$ $(j \geq 1)$ are the sequence of i.i.d. vectors with $\mathcal{E}[\mathbf{e}_j^{(x)}] = \mathbf{0}$ and $\mathcal{E}[\mathbf{e}_j^{(x)} \mathbf{e}_j^{(x)\prime}] = \boldsymbol{\Sigma}_x$. (We use $\|\mathbf{C}_j^{(x)}\| = \max_{k,l=1,\dots,p} |c_j^{(x)}(k,l)|$ for $c_j^{(x)}(k,l)$ is the (k,l)-th element of $\mathbf{C}_j$ and $\sum_{j=0}^{\infty} \|\mathbf{C}_j\| < +\infty$.)

Then we can represent the spectral density $p \times p$ matrix of $\Delta\mathbf{x}_i$ as

$$f_{\Delta x}(\lambda) = \frac{1}{\pi} \left(\sum_{j=0}^{\infty} \mathbf{C}_j^{(x)} \mathrm{e}^{2i\lambda j} \right) \boldsymbol{\Sigma}_x \left(\sum_{j=0}^{\infty} \mathbf{C}_j^{(x)\prime} \mathrm{e}^{-2i\lambda j} \right) \quad \left(-\frac{\pi}{2} \leq \lambda \leq \frac{\pi}{2} \right), \tag{41}$$

where $i^2 = -1$, λ is the frequency, and $\boldsymbol{\Sigma}_x$ is the variance–covariance matrix of $\mathbf{e}_j^{(x)}$ (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

$$f_{\Delta y}(\lambda) = f_{\Delta x}(\lambda) + \frac{1}{\pi}(1 - e^{2i\lambda})\boldsymbol{\Sigma}_v(1 - e^{-2i\lambda}). \tag{42}$$

Since the transformed random vectors $z_k \ (k = 1, \dots, n)$ correspond to the Fourier transformed vectors of $\Delta\mathbf{y}_i \ (i = 1, \dots, n)$ except the initial condition $\mathbf{y}_0$, it is possible to estimate the spectral density matrix $f_{\Delta x}(\lambda)$ at the zero-frequency from the sequence of observations of $\mathbf{z}_k \ (k = 1, \dots, n)$ and we write

$$\boldsymbol{\Omega}_x = \pi f_{\Delta x}(0) = \pi f_{\Delta y}(0). \tag{43}$$

We denote the components of $\boldsymbol{\Omega}_x = (\Omega_{x.ij}) \ (i, j = 1, 2)$ and we have the same form of SIML estimate in (11) except the fact that the estimated parameter is $\boldsymbol{\Omega}_x$ instead of $\boldsymbol{\Sigma}_x$. Then the asymptotic results stated in Theorem 4.3 hold when we have (1), (2), and (40) with mild regularity conditions using straight-forward, but lengthy arguments. It is because the SIML estimation can be regarded a kernel estimation of the spectral density matrix at zero frequency.

5.2 Reduced rank condition

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

$$\text{rank}[\boldsymbol{\Sigma}_x] = q_x, \quad 1 \le q_x < p, \tag{44}$$

where we can represent $\boldsymbol{\Sigma}_x = \mathbf{B}\mathbf{B}'$ and $\mathbf{B}$ is a $p \times q_x$ matrix (its rank is q_x).

Then the example in Sect. 3.2 corresponds to the case when $q_x = 1 < p$ and $p \ge 2$.

In the more general cases, however, there is a parametrization problem for the $p \times p$ matrix $\boldsymbol{\Sigma}_x$, whose rank is $q_x \ (1 \le q_x < p, p \ge 2)$. Then we can take a $p \times r_x \ (r_x = p - q_x)$ matrix $\boldsymbol{\beta}$ such that $\mathbf{B}'\boldsymbol{\beta} = \mathbf{O}$ and a normalization such as $\boldsymbol{\beta}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\beta} = (\text{diag } c_{ii})$ for instance. The algebra of Sect. 3.2 can be extended by using the matrix formulae such that for a positive definite matrix $\mathbf{A}$, a $p \times q_x$ matrix $\mathbf{B}$ (the rank is q_x) and a $p \times r_x$ matrix $\boldsymbol{\beta}$ (the rank is r_x), we have $\boldsymbol{\Sigma}_x = \mathbf{B}\mathbf{B}'$, $|\mathbf{A} + \mathbf{B}\mathbf{B}'| = |\mathbf{A}||\mathbf{I}_{q_x} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}|$ and

$$[\mathbf{A} + \mathbf{B}\mathbf{B}']^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}[\mathbf{I}_{q_x} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}]^{-1}\mathbf{B}'\mathbf{A}^{-1}. \tag{45}$$

Hence we can use the characteristic vectors with the r_x smaller characteristic roots of the equation [which is similar to (30)].

The special case when $q_x = p - 1$ has attracted special attention among econometricians or economists because there exists a unique vector $\boldsymbol{\beta}$ such that $\boldsymbol{\Sigma}_x\boldsymbol{\beta} = \mathbf{0}$ with a normalization. It has been called the single structural equation in

traditional econometrics or the co-integrated relation in time series econometrics and we can extend the developments of Sect. 3.2. When $p > 2$ and $0 < q_x < p - 1$, there are $r_x (= p - q_x)$ co-integrated vectors satisfying $\mathbf{B}'\boldsymbol{\beta} = \mathbf{O}$ and then there is an identification problem on the vectors consisting of $\boldsymbol{\beta}$.

Remark 5.1 The problem of reduced rank condition was a central issue in Anderson (1984) for the case of independent observations in statistical multivariate analysis. For the non-stationary time series without measurement-errors, Engle and Granger (1987), and Johansen (1995) have developed the statistical inference and they called the resulting equations from the non-stationary multiple time series as co-integrated relations. It is possible to develop the testing procedure of the rank condition for hidden non-stationary components.

5.3 Higher order integrated processes

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

$$\Delta^d \mathbf{x}_i = \mathbf{v}_i^{(x)}, \tag{46}$$

$\mathcal{E}[\mathbf{v}_i^{(x)}] = \mathbf{0}$, and $\mathcal{E}[\mathbf{v}_i^{(x)}\mathbf{v}_i^{(x)'}] = \boldsymbol{\Sigma}_x$.

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

$$\mathbf{Z}_n^{(2)} = \left(\mathbf{z}_k^{(2)'}\right) = \mathbf{K}_n(\mathbf{Y}_n - \bar{\mathbf{Y}}_0), \quad \mathbf{K}_n = \mathbf{P}_n \mathbf{C}_n^{-2}. \tag{47}$$

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

$$\hat{\boldsymbol{\Sigma}}_{x,\text{SIML}} = \frac{1}{m_n} \sum_{k=1}^{m_n} \mathbf{z}_k^{(2)} \mathbf{z}_k^{(2)'}. \tag{48}$$

We prepare the next lemma.

Lemma 5.1 *Let*

$$\mathbf{K}_n = (b_{ij}^{(2)}) = \mathbf{P}_n \mathbf{C}_n^{-2}. \tag{49}$$

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

$$\sum_{j=1}^{n} b_{ij}^{(2)} b_{i'j}^{(2)} = \delta(i, i')\left[2\sin\left(\frac{\pi}{2}\frac{2i-1}{2n+1}\right)\right]^4 + O\left(\frac{1}{n}\right). \tag{50}$$

Using the above lemma, it is straight-forward to obtain the next result for the case of $d \geq 1$. We omit the proof because it is essentially the same to the case when $d = 1$ except Lemma 5.1. It is clear that the result holds for a positive integer d.

Theorem 5.2 *Assume $p \geq 1$, $d = 2$ and $m_n/n \longrightarrow 0$ as $n \longrightarrow \infty$, where $m_n = [n^\alpha]$ $(0 < \alpha < 1)$. Under the assumption of existence of fourth order moments, the SIML estimator of Σ_x is consistent with $0 < \alpha < 1$ and asymptotically normal with $0 < \alpha < 0.8$ as $n \longrightarrow \infty$.*

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

$$a_{kn}^{(2)} = \left[2 \sin \frac{\pi}{2n+1} \left(k - \frac{1}{2} \right) \right]^4 \tag{51}$$

in the present case and we need the corresponding bias correction for estimating the variance–covariance matrix Σ_v.

***Remark* 5.2** Akaike (1989) and Kitagawa (2010) have proposed to use the ML estimation with some restriction for the filtering problem of the same models when $d \geq 1$. The SIML estimation would be a useful tool for the state space modeling of non-stationary multivariate time series because it does not have any computational problem without any restriction and it has an asymptotic robustness. The SIML method may give reasonable estimates not only for the coefficients parameters, but also the variance-covariance matrices even if we do not know $d = 1, 2$ by taking m_n appropriately.

6 Concluding remarks

In this study, we introduced the non-stationary errors-in-variables models. We first illustrated the reason why the presence of noise term in the non-stationary time series, even if it was small, forces to change the standard way of thinking in time series analysis. Then, we discussed the finite sample and large sample properties of the ML and SIML estimation methods for the non-stationary errors-in-variables models when there are non-stationary trends and noise components (or the measurement errors). We have found that the Gaussian likelihood function shows non-concave shape in some cases and the ML method does work when the Gaussianities of non-stationary and stationary components hold with some restriction such as the signal–noise variance ratio in the parameter space. The SIML estimation has the asymptotic robust properties under general conditions of existence of fourth-order moments for consistency and asymptotic normality. We have investigated the conditions for the asymptotic properties of the ML and SIML estimations as well as some simulations. The SIML method gives reasonable estimates for the coefficients parameters when the random variables do not necessarily follow the Gaussian distribution and we do not have much knowledge

on the value of variance ratio c by taking m_n appropriately. Since we usually do not have information on the magnitude of the signal–noise variance ratio and the precise distribution of noises in advance when we observe non-stationary data, it is important to use the statistical method which does not depend on these conditions for practical purpose.

There are several possible extensions and some discussions are given in Sects. 3 and 5. It may be interesting to see if the results reported in this paper would be valid when there are non-stationary seasonal components. There can be many applications of the methods we have discussed because many economists are interested in macro-economic data and their relations using seasonally adjusted data. Another issue may be that some econometricians have relied on the large sample asymptotic theory and used the approximations based on Brownian motions, which are the limits of random walks. Although there are some procedures based on Quadratic Variation [see (12)] of the observed data to estimate the variance–covariances of innovations and the long-run variance–covariances, there is a fundamental problem as illustrated by examples in Sect. 2.2.

If we ignore the presence of noise components and/or measurement components, there could be serious problems. For instance, the actual sample size of macro-economic data is usually not large and often we have measurement errors. In such situations we should be careful to use such asymptotic theory and we need to investigate the finite sample properties of alternative estimation methods and their improvements. The precise consequences of these effects are currently under investigation and some progress on the filtering and smoothing problem of noisy non-stationary multivariate time series has been reported in Kunitomo and Sato (2019).

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

$$\mathcal{E}[z_{ik} z_{jk} z_{i'k} z_{j'k}] = q_{ij} q_{i'j'} + q_{ii'} q_{jj'} + q_{ij'} q_{i'j}, \tag{A.52}$$

where $\mathbf{Q} = (q_{ij})$. Then it is straight-forward to obtain the results.   □

Proof of Theorem 4.2 Using Lemma 4.1, it is possible to obtain the variance of L^*_{12n}, which converges to 0 as $n \longrightarrow \infty$. We set

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

and

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

Then

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

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

$$
\begin{aligned}
&= \mathcal{E}\left[\sum_{k=1}^{n} \mathrm{tr}\mathbf{A}_k(\mathbf{z_k}\mathbf{z}_k' - \mathbf{Q}_0) \right]^2 \\
&= \mathcal{E}\left[\sum_{k,k'=1}^{n} \mathrm{tr}\mathbf{A}_k(\mathbf{z}_k\mathbf{z}_k' - \mathbf{Q}_0)\mathbf{A}_{k'}(\mathbf{z}_{k'}\mathbf{z}_{k'}' - \mathbf{Q}_0) \right] \\
&= \left[\sum_{k=1}^{n} \mathcal{E}\left[(\mathbf{z}_k'\mathbf{A}_k\mathbf{z}_k)^2 - (\mathrm{tr}\mathbf{A}_k\mathbf{Q}_0)^2\right] \right] \\
&= \sum_{k=1}^{n} 2\mathrm{tr}(\mathbf{A}_k\mathbf{Q}_0\mathbf{A}_k\mathbf{Q}_0). \quad\quad\quad\quad \text{(A.53)}
\end{aligned}
$$

If $\mathbf{Q} = \mathbf{Q}_0$, then it is $2pn$.

Next, we investigate the first term and last term of L_{1n}^* in detail. After simple algebra, L_{13n}^* can be re-expressed as

$$
\begin{aligned}
L_{13n}^* =& \operatorname{tr}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}(\theta_0) + \frac{1}{n}\sum_{k=1}^n a_{kn}^{*-1}\mathbf{b}(\theta_0')\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0) \\
&- \frac{1}{n}\sum_{k=1}^n \frac{1}{a_{kn}^* + c}\mathbf{b}'[\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}(\theta)\boldsymbol{\Sigma}_v^{-1} + a_{kn}^{*-1}\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0)\mathbf{b}(\theta_0)'\boldsymbol{\Sigma}_v^{-1}]\mathbf{b} \\
=& \operatorname{tr}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}(\theta_0) \\
&+ \frac{1}{n}\sum_{k=1}^n \frac{1}{a_{kn}^*(a_{kn}^* + c)}\Big[(a_{kn}^* + c)\mathbf{b}(\theta_0)'\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0) - a_{kn}^*\mathbf{b}'\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}_v(\theta_0)\boldsymbol{\Sigma}_v^{-1}\mathbf{b} \\
&- (\mathbf{b}'\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta))^2\Big] \\
=& \operatorname{tr}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}(\theta_0) \\
&+ \frac{1}{n}\sum_{k=1}^n \frac{1}{a_{kn}^*(a_{kn}^* + c)}\Big[a_{kn}^*\mathbf{b}(\theta_0)'\boldsymbol{\Sigma}_v^{-1}(\boldsymbol{\Sigma}_v - \boldsymbol{\Sigma}_v(\theta_0))\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0) \\
&+ \mathbf{b}(\theta_0)'\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0)\mathbf{b}'\boldsymbol{\Sigma}_v^{-1}\mathbf{b} - (\mathbf{b}'\boldsymbol{\Sigma}_v^{-1}\mathbf{b}(\theta_0))^2\Big]
\end{aligned}
$$

because $c = \mathbf{b}'\boldsymbol{\Sigma}_v^{-1}\mathbf{b}$.

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 each term in L_{1n}^*, which are of the same order, as

$$
\frac{1}{n}\sum_{k=1}^n \log|a_{kn}^*\boldsymbol{\Sigma}_v| + \frac{1}{n}\sum_{k=1}^n \log[1 + a_{kn}^{*-1}\mathbf{b}'\boldsymbol{\Sigma}_v^{-1}\mathbf{b}] \tag{A.54}
$$

and

$$
\operatorname{tr}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}_v(\theta_0) + \frac{1}{n}\sum_{k=1}^n \frac{\mathbf{b}'\boldsymbol{\Sigma}_v^{-1}(\boldsymbol{\Sigma}_v - \boldsymbol{\Sigma}_v(\theta_0))\boldsymbol{\Sigma}_v^{-1}\mathbf{b}}{a_{kn}^* + \mathbf{b}'\boldsymbol{\Sigma}_v^{-1}\mathbf{b}}. \tag{A.55}
$$

Using the inversion formula and its determinant in (27) and (28), the sum of above two terms can be written as

$$
\frac{1}{n}\sum_{k=1}^n \log|a_{kn}^*\boldsymbol{\Sigma}_v + \mathbf{bb}'| + \frac{1}{n}\sum_{k=1}^n \operatorname{tr}[a_{kn}^*\boldsymbol{\Sigma}_v + \mathbf{b}'\mathbf{b}]^{-1}[a_{kn}^*\boldsymbol{\Sigma}_v(\theta_0) + \mathbf{bb}'].
$$

Using Lemma 3.2.2 of Anderson (2003), the minimum of each terms occurs at $a_{kn}^*\boldsymbol{\Sigma}_v + \mathbf{bb}' = a_{kn}^*\boldsymbol{\Sigma}_v(\theta_0) + \mathbf{bb}'$, that is, $\boldsymbol{\Sigma}_v = \boldsymbol{\Sigma}_v(\theta_0)$ under the assumption of

positive definiteness. Hence the global minimum of the (minus) log-likelihood function occurs iff $\mathbf{b} = \mathbf{b}(\theta_0)$ and $\boldsymbol{\Sigma}_v = \boldsymbol{\Sigma}_v(\theta_0)$.

The rest of our arguments for the consistency follow from the general arguments for the extremum estimation [see Theorem 4.1.1 of Amemiya (1985), for instance] and we have the desired result. □

Proof of Lemma 5.1 We set

$$\theta_{kj} = \frac{2\pi}{2n+1}\left(j - \frac{1}{2}\right)\left(k - \frac{1}{2}\right), \quad \theta_k = \frac{2\pi}{2n+1}\left(k - \frac{1}{2}\right), \tag{A.56}$$

and

$$b_{kj}^{(2)} = \frac{1}{\sqrt{2n+1}}[(1 - e^{i\theta_k})^2 e^{i\theta_{kj}} + (1 - e^{-i\theta_k})^2 e^{-i\theta_{kj}}] \tag{A.57}$$

for $j, k = 1, \ldots, n$.

Then using direct evaluations

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

Also for $k \neq k'$ we find that

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

Then after some algebra, we have

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

Hence, we finally find that

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

and then it is proportional to

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

when $k/n \longrightarrow 0$ as $n \to \infty$. $\square$

References

Akaike, H. (1989). Choosing priors and its applications. In Y. Suzuki & N. Kunitomo (Eds.), *Bayesian statistics and applications*. Tokyo: Tokyo University Press **(in Japanese)**.

Amemiya, T. (1985). *Advanced econometrics*. Oxford: Blackwell.

Anderson, T. W. (1971). *The statistical analysis of time series*. New York: Wiley.

Anderson, T. W. (1984). Estimating linear statistical relationships. *Annals of Statistics*, *12*, 1–45.

Anderson, T. W. (2003). *An introduction to statistical multivariate analysis* (3rd ed.). New York: Wiley.

Anderson, T. W., & Takemura, A. (1986). Why do non-invertible moving average occur? *Journal of Time Series Analysis*, *7–4*, 235–254.

Anderson, T. W., Kunitomo, N., & Morimune, K. (1986). Comparing single-equation estimators in a simultaneous equation system. *Econometric Theory*, *2*, 1–32.

Andrews, D. K. (1999). Estimation with a parameter is on a boundary. *Econometrica*, *67*, 1341–1383.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis, forecasting and control*. San Francisco: Holden-Day.

Chang, Y., Jiang, B., & Park, J. (2008). *Using Kalman filter to extract and test for common stochastic trends*. Unpublished manuscript.

Engle, R., & Granger, C. W. J. (1987). Co-integration and error correction. *Econometrica*, *55*, 251–276.

Fuller, W. (1987). *Measurement error models*. New York: Wiley.

Granger, C. W. J., & Newbold, P. (1977). *Forecasting economic time series*. Cambridge: Academic Press.

Hayashi, F. (2000). *Econometrics*. Princeton: Princeton University Press.

Johansen, S. (1995). *Likelihood based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.

Kitagawa, G. (2010). *Introduction to time series modeling*. Boca Raton: CRC Press.

Kunitomo, N., & Sato, S. (2017). *Trend, seasonality and economic time series: The nonstationary errors-in-variables models*. MIMS-REB-SDS-9, MIMS, Meiji University. https://mims.meiji.ac.jp/publications/datascience-e.html.

Kunitomo, N., & Sato, S. (2019). *A robust-filtering method for noisy non-stationary time series with an application to Japanese macro-consumption*. Unpublished manuscript. MIMS-RBP-SDS-9, MIMS, Meiji University

Kunitomo, N., Sato, S., & Kurisu, D. (2018). *Separating information maximum likelihood method for high-frequency financial data*. Berlin: Springer.

Morgenstern, O. (1950). *On the accuracy of economic observations, (1966, revised edition)*. Princeton: Princeton University Press.

Nerlove, M., Grether, D. M., & Carvalho, J. L. (1995). *Analysis of economic time series: A synthesis* (Revised ed.). New York: Academic Press.

Rothenberg, T. (1971). Identification in parametric models. *Econometrica*, *39*, 577–591.