11-617

no. 01-13E

c. 2

Statistics Canada Statistique Canada

Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes-auprès des entreprises

Canada

SURVEY ERROR MODELLING IN THE PRESENCE OF BENCHMARKS

By

Zhao-Guo Chen[1] and Ka Ho Wu[2]

BSMD-2001-013E

STATISTICS CANADA STATISTIQUE CANADA

NOV 30 2001

LIBRARY BIBLIOTHÈQUE

SURVEY ERROR MODELLING IN THE PRESENCE OF BENCHMARKS

by

Zhao-Guo Chen[1] and Ka Ho Wu[2]

November 2001
Time Series Research and Analysis Centre
Business Survey Methods Division
Statistics Canada

[1]Time Series Centre, 3H Coats Building, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6 (zhao-guo.chen@statcan.ca)
[2]Department of Statistics The Chinese University of Hong Kong, Shatin, Hong Kong (wu@sta.cuhk.edu.hk)

ABSTRACT

Data for a socio-economic variable obtained from a repeated survey contain sampling error. Usually estimates of the variance of the error are obtained in the survey process and published; but estimates of the autocorrelation of the error series (equivalently, a fitted time series model) are rarely given. Knowing autocorrelation of the survey error series, some advanced approaches for predicting the variable can work. This paper proposes a method of modelling monthly survey error using annual benchmarks as additional information under a very general model assumption for the variable. The fitted model is thus more data-based than those obtained from secondary analysis (e.g. Scott, Smith and Jones, 1977) where only monthly data are used.

Résumé

Les données au sujet d'une variable socio-économique provenant d'enquêtes répétées contiennent de l'erreur. En général des estimations de la variance de l'erreur s'obtiennent du processus-même de l'enquête et font l'objet de publication. Par contre les estimations de l'autocorrélation de l'erreur (c'est-à-dire du modèle suivi par l'erreur) sont rarement disponibles. Or il est reconnu que l'autocorrélation de l'erreur d'échantillonnage est très utile pour l'amélioration des prévisions de la variable. Ce document propose une méthode de modélisation de l'erreur d'échantillonnage en présence de jalons annuels, basée sur des hypothèses très générales au sujet de la variable d'intérêt. Le modèle ajusté est donc plus dicté par les /données disponibles que le modèle qui serrait obtenu par la méthode de l'analyse secondaire (par exemple, Scott, Smith et Jones, 1977) où seulement les données mensuelles sont utilisées.

1. Introduction

Data for a socio-economic variable obtained from repeated survey contains sampling error. Usually, estimates of the variance of the error are obtained in the survey process and published by statistical agencies; but estimates of the autocorrelation of the error series (equivalently, a fitted time series model) are rarely given. Knowing the autocorrelation, some advanced methods, such as signal extraction, can be used to predict the variable of interest. That may drastically reduce the prediction error.

The autocorrelation of survey error heavily depends on the way of the same sampling units appear in the samples more than one occasion, the so-called overlap. If values of individual units, or other elementary estimates such as values for rotating panels, are available, methods of modelling survey error or estimating the autocorrelation of survey error are referred to as *primary analysis* of survey error (e.g. Scott, Smith and Jones, 1977; Pfeffermann, 1991).

However, in most situations, values of neither individual units nor panels are available due to various reasons including the need for confidentiality, the problem of inadequate data identification and the lack of suitable program to handle repeated survey. In such cases, methods of modelling survey error or estimating autocorrelation of survey error have to rely on the aggregated data in the published form. These methods are then referred to as *secondary analysis* of survey error (e.g. Scott, Smith and Jones, 1977; Pfeffermann, 1991).

Currently, these methods of modelling survey error fit an ARIMA model to the data and decompose the resulting model into two component models. Unfortunately, the conditions for such a decomposition are very restrictive which, combined with the lack of auxiliary information, force the analysts to be quite subjective. For example, in Scott, Smith and Jones (1977), the AR parts of the models for the observations, the variable of interest and the survey error are assumed to be identical. One of our

efforts is to relax model assumptions on the variable of interest in order to make the model as general as possible and no parameter specification is needed.

In the literature, methods of using auxiliary information for improving the prediction of variable are very common. A situation called benchmarking (see (1.1) and (1.2) below), where the available information takes the form of additional estimates related to the same variable of interest but observed much less frequently and much more precisely. For examples, see Denton (1971), Hillmer and Trabelsi (1987), Chollete and Dagum (1994), Durbin and Quenneville (1997), Chen, Cholette and Dagum (1997). Simulation examples in Chen, Chollette and Dagum (1997) show that the prediction error of the variable of interest can be substantially reduced when some advanced benchmarking methods are applied. However, these methods require the autocorrelation of the survey error or its model.

In this paper, we propose a method of modelling survey error by using the aggregated survey data as well as available benchmarks (additional information). We denote this kind of approaches as the *extended secondary analysis* of survey error.

The general set-up is as follows. We have monthly survey data $y(t)$ which are the sum of the variable $\eta(t)$ and survey error $e(t)$ with zero mean, i.e.

$$y(t) = \eta(t) + e(t), \quad t = 1, \cdots, n, \quad E\ e(t) = 0. \tag{1.1}$$

We also have the annual sum $z(T)$, called benchmarks, obtained from a more reliable source, i.e.

$$z(T) = \sum_{t \in T} \eta(t) + \psi(T), \quad T = 1, \cdots, N, \quad n = 12N, \tag{1.2}$$

where the notation $t \in T$ means month t is in year T. Methods or procedures of predicting $\eta(t)$ using both the monthly survey data $y(t)$ and the annual benchmarks $z(T)$ are referred to as benchmarking and the predictions of $\eta(t)$ are called the benchmarked values or the benchmarking predictions. $\psi(T)$ in (1.2) is the error; its variance is usually much smaller than the variance of $e(t)$. In most practical situations, $\psi(T)$

can be ignored (i.e. $\psi(T) = 0$). In this case, $z(T)$ are called binding benchmarks; otherwise non-binding benchmarks. In this work, we only consider the case of binding benchmarks.

To be more focus, we restrict ourselves to the case where the survey error $e(t)$ follows a first order autoregressive (AR(1)) model

$$e(t) = \phi\, e(t-1) + \xi(t), \tag{1.3}$$

where $0 \leq \phi < 1$ and $\{\xi(t)\}$ is a white noise series with mean zero and variance σ^2. When the survey design has no complicated rotation (such as a penal may re-enter the sample), this assumption is reasonable. See for example Scott, Smith and Jones (1977). Furthermore, the model suggested by Lee (1990) resemble to an AR(1) model. Perhaps, our proposed method can be extended to deal with more complicated models such as those in Bell and Wilcox (1993), Pfeffermann and Scott (1997). However, adding more parameters in the model generates substantial difficulties in the small sample environments. The problem we try to solve is how to use $\{y(t),\ t = 1, \cdots, n\}$ and $\{z(T),\ T = 1, \cdots, N\}$ to estimate ϕ and σ^2 in (1.3). We only illustrate the case of monthly survey data with annual benchmarks. The principle for quarterly survey data with annual benchmarks, or monthly survey data with quarterly benchmarks are similar.

The layout of the paper is as follows. In Section 2, further assumptions for model (1.1) are introduced and some basic estimation formulae are derived. In Section 3, we show that in practice, if the number of available benchmarks is small, the formulae in Section 2 may not provide reasonable estimates. To overcome the difficulties, we propose some alternative strategies. To illustrate the superiority of our procedure, a simulation study is given in Section 4. Concluding remarks are provided in Section 5. Technical details are left to the Appendices.

2. Model assumptions and estimation formulae

2.1 Benchmarking model and autocovariances of survey error

The vector representation of (1.1) and (1.2) for binding benchmarks can be written as

$$\begin{cases} \mathbf{y} &= \boldsymbol{\eta} + \mathbf{e} \\ \mathbf{z} &= \mathbf{L}\boldsymbol{\eta} \end{cases} \tag{2.1}$$

where $\mathbf{y} = (y(1), \cdots, y(n))', \mathbf{z} = (z(1), \cdots, z(N))'$ etc., and

$$\mathbf{L} = \begin{pmatrix} \mathbf{1}' & .. & 0 \\ . & .. & . \\ 0 & .. & \mathbf{1}' \end{pmatrix}_{N \times n}, \quad \mathbf{1}' = (1, \cdots, 1)_{1 \times 12}, \quad n = 12N.$$

If the covariance matrix $\mathbf{V_e}$ of $\mathbf{e}$ is known, taking benchmarking via regression as example, the benchmarked value for $\boldsymbol{\eta}$ is (see Chollette and Dagum, 1994)

$$\hat{\boldsymbol{\eta}} = \mathbf{y} + \mathbf{V_e}\mathbf{L}'(\mathbf{L}\mathbf{V_e}\mathbf{L}')^{-1}(\mathbf{z} - \mathbf{L}\mathbf{y}), \tag{2.2}$$

where

$$\mathbf{V_e} = \begin{pmatrix} v_e(0) & .. & v_e(n-1) \\ . & .. & . \\ v_e(n-1) & .. & v_e(0) \end{pmatrix}. \tag{2.3}$$

For the special case of model (1.3),

$$v_e(k) = \sigma^2\phi^k/(1-\phi^2), \quad k = 0, 1, 2, \cdots. \tag{2.4}$$

Sometimes, survey experts may provide a statistical model for the survey error $e(t)$ or directly provide values of the autocovariance function $v_e(k)$, i.e. $\mathbf{V_e}$ is known. However, in most practical situations, $\mathbf{V_e}$ is unknown and only an estimate of $v_e(0)$ is given. In this paper, we assume that $e(t)$ follows the AR(1) model (1.3) and the problem is that of estimating ϕ and σ^2 from $\mathbf{y}$ and $\mathbf{z}$. Thus the estimates of $v_e(k)$, $k = 1, 2, ...$, are obtained from (2.4) and consequently $\hat{\boldsymbol{\eta}}$ can be calculated by (2.2).

2.2 Original sketch for estimating $v_e(k)$ using benchmarks

Let $\varepsilon(T) = e(12T-11) + \cdots + e(12T)$, the sum of monthly survey error in year T, and $\boldsymbol{\varepsilon} = (\varepsilon(1), \cdots, \varepsilon(N))'$. It can be obtained from $\mathbf{y}$ and $\mathbf{z}$, i.e.

$$\boldsymbol{\varepsilon} = \mathbf{Le} = \mathbf{L}(\mathbf{y} - \boldsymbol{\eta}) = \mathbf{Ly} - \mathbf{z}. \tag{2.5}$$

From (2.5), we have

$$\mathbf{V}_{\boldsymbol{\varepsilon}} = Var(\boldsymbol{\varepsilon}) = \mathbf{LV_eL}' \text{ and } \mathbf{V}_{\mathbf{e}\boldsymbol{\varepsilon}} = Cov(\mathbf{e}, \boldsymbol{\varepsilon}) = \mathbf{V_e}\,\mathbf{L}'. \tag{2.6}$$

It is difficult to estimate $\mathbf{V_e}$ by using the first equation (from $\mathbf{V}_{\boldsymbol{\varepsilon}}$), but we can estimate $\mathbf{V_e}$ by using the second equation (from $\mathbf{V}_{\mathbf{e}\boldsymbol{\varepsilon}}$). The first K elements of the first column of $\mathbf{V}_{\mathbf{e}\boldsymbol{\varepsilon}}$, denoted by $v_{e\varepsilon}(k,1)$, $k = 1, 2, \cdots, K < n$, can be expressed as

$$(v_{e\varepsilon}(1,1), \cdots, v_{e\varepsilon}(K,1))' = \mathbf{F}(v_e(0), \cdots, v_e(K-1))', \tag{2.7}$$

where the $K \times K$ matrix $\mathbf{F}$ has elements 0, 1 or 2. Knowing the values of $v_{e\varepsilon}(k,1)$, $k = 1, \cdots, K$, the values of $v_e(k), k = 0, \cdots, K-1$, can be obtained from (2.7). For $K > 2$, if there is an estimate $\hat{v}_{e\varepsilon}(k,1)$ of $v_{e\varepsilon}(k,1)$, we may replace $v_{e\varepsilon}(k,1)$ by $\hat{v}_{e\varepsilon}(k,1)$ on the left hand side of (2.7) and expressed $v_e(k)$ as (2.4) on the right hand side of (2.7) to form a non-linear regression model so that the least squares estimates of ϕ and σ^2 can be obtained.

2.3 Assumptions for $\eta(t)$

Case 1. Assume that the series $\eta(t)$ is stationary with $E[\eta(t)] = \mu$ (without loss of generality, let $\mu = 0$) and $\mathbf{e}$ and $\boldsymbol{\eta}$ in (2.1) are uncorrelated (so does $\boldsymbol{\varepsilon}$ and $\boldsymbol{\eta}$), then

$$E(\mathbf{y}\,\boldsymbol{\varepsilon}') = E[(\boldsymbol{\eta} + \mathbf{e})\boldsymbol{\varepsilon}'] = E(\boldsymbol{\eta}\boldsymbol{\varepsilon}') + E(\mathbf{e}\,\boldsymbol{\varepsilon}') = \mathbf{V}_{\mathbf{e}\boldsymbol{\varepsilon}}. \tag{2.8}$$

We wish to use the sample covariances of $e(t)$ and $\varepsilon(T)$ to estimate $v_{e\varepsilon}(k,1)$ [c.f. (2.21)]. $\varepsilon(1), \cdots, \varepsilon(N)$ can be obtained from $\mathbf{y}$ and $\mathbf{z}$ by (2.5), but none of the $e(t)$ is given. According to (2.8), $\mathbf{e}$ can be replaced by $\mathbf{y}$ because the $\boldsymbol{\eta}$ part in the sample covariances of $y(t)$ and $\varepsilon(T)$ for estimating $E(\boldsymbol{\eta}\boldsymbol{\varepsilon}') = 0$ has only a little effect when

N is large. For small N, this effect can be examined by simulation. Hence $\mathbf{V}_{e\varepsilon}$ can be estimated by $\mathbf{y}$ and $\boldsymbol{\varepsilon}$.

Case 2. $\eta(t)$ are fixed numbers (Cholette and Dagum, 1994). Although $E(\boldsymbol{\eta}\boldsymbol{\varepsilon}') = 0$ holds, it is not justified to replace the unknown $e(t)$ by $y(t)$ for estimating $E(\mathbf{e}\ \boldsymbol{\varepsilon}') = \mathbf{V}_{e\varepsilon}$. The major reason is that there is no definition for the theoretical covariance between $\eta(t)$ and $\varepsilon(T)$ and no existing theory to support the idea that the "sample covariance" can be ignored. Hence, this case will not be considered in this article.

Case 3. Most socio-economic series do not stay at the same level, they usually drift from one level to another constantly. Therefore, it is more interesting to consider the following very general nonstationary model

$$\nabla\nabla_{12}\eta(t) = \eta^*(t), \tag{2.9}$$

where $\eta^*(t)$ is a zero mean stationary series, $\nabla = 1 - B$, $\nabla_{12} = 1 - B^{12}$ and B is the backshift operator defined by $B^k\eta(t) = \eta(t-k)$. Model (2.9) can be used to fit many real series and have been widely used. In (2.9), $\eta(t)$ may be called difference stationary (denoted by DS, Nelson and Plosser, 1982). In this paper, $\eta^*(t)$ is allowed to be over-differenced.

Case 4. $\eta(t)$ is the sum of a trend (fixed numbers) and a stationary series. Such a nonstationary series is called trend stationary (denoted by TS, Nelson and Plosser, 1982) which is quite useful in many practical circumstances. For general TS models, we may also face the similar difficulties as in Case 2. However, we may consider a special and the most useful TS model,

$$\eta(t) = \text{linear trend} + \text{seasonality} + \zeta(t), \tag{2.10}$$

where $\zeta(t)$ is stationary and mean zero. Since the linear trend and the seasonality disappear after $\nabla\nabla_{12}$ is operated, $\eta(t)$ also follows (2.9).

2.4 Estimating $v_e(k)$ under nonstationary assumptions for $\eta(t)$

The nonstationary model (2.9), Case 3, is very general. It may also cover (2.10) and Case 1, but notice that $\eta^*(t) = \nabla\nabla_{12}\eta(t)$ (for (2.10), $\eta^*(t) = \nabla\nabla_{12}\zeta(t)$) is then an over-differenced series. We now derive a relationship similar to (2.7) which is the basic formula for estimating $v_e(k)$ under the nonstationary assumptions.

From (1.1) and (2.9), we have

$$y^*(t) = \nabla\nabla_{12}y(t) = \eta^*(t) + e^*(t), \quad t = 14, \cdots, n, \tag{2.11}$$

where

$$e^*(t) = e(t) - e(t-1) - e(t-12) + e(t-13). \tag{2.12}$$

Now, all three variables in (2.11) are stationary with mean 0. In addition, $\eta^*(t)$ and $e^*(t)$ are mutually uncorrelated.

Put $v_{e^*\varepsilon}(k) = v_{e^*\varepsilon}(k,1) = E[e^*(k)\varepsilon(1)]$, where $k = 14, 15, \cdots$, we have

$$(v_{e^*\varepsilon}(14), \cdots, v_{e^*\varepsilon}(K))' = \mathbf{F}_K(v_e(0), \cdots, v_e(K-1)), \tag{2.13}$$

where the $(K-13) \times K$ matrix $\mathbf{F}_K$ is truncated from the following matrix

	(0)	(1)	. .	(5)	(6)	(7)	. . .	(11)	(12)	(13)	. . .
(14)		-2						1		1	
.			.				.				.
.				.				.			.
.					.		.				.
(18)				-2		1					
(19)					-1						
(20)				1		-2					
.			.				.				
.				.				.			
.			.					.			
(24)		1						-2			
(25)	1								-2		
(26)		1								-2	
.			.								.
.				.							.

The missing elements in the matrix are 0; the numbers in the parenthesis on the top of the matrix correspond to k in $v_e(k)$, $k = 0, 1, \ldots$; the numbers in the parenthesis on the left of the matrix correspond to k in $v_{e^*\varepsilon}(k)$, $k = 14, 15, \ldots$.

With $e(t)$ following the AR(1) model (1.3), $v_e(k)$ in (2.13) should be replaced by (2.4). Combining this with the expression of $\mathbf{F}_K$, we have

$$v_{e^*\varepsilon}(k) = v_{e^*\varepsilon}(k, \phi, \sigma^2) = \rho(k, \phi)\sigma^2/(1 - \phi^2), \quad k = 14, 15, \cdots, \tag{2.14}$$

where

$$\rho(k, \phi) = -2\phi^{k-13} + \phi^{25-k} + \phi^{k-1}, \quad k = 14, 15, \cdots. \tag{2.15}$$

Figure 2.1 displays $\rho(k, \phi)$ with $k = 14, 15, \ldots, 25$ for various values of ϕ (indicated beside the lines; where the dotted lines correspond to $\phi = 0.9, 0.95$ and 0.99).

Now, if we have estimates $\hat{v}_{e^*\varepsilon}(k)$ of $v_{e^*\varepsilon}(k)$, a nonlinear regression model with error $\nu(k)$,

$$\hat{v}_{e^*\varepsilon}(k) = \rho(k, \phi)\sigma^2/(1 - \phi^2) + \nu(k), \quad k = 14, \cdots, K, \tag{2.16}$$

can be used for estimating ϕ and σ^2. Denote (ϕ_1, σ_1^2) as the least squares estimate obtained by minimizing

$$S(\phi, \sigma^2) = \sum_{k=14}^{K} \{\hat{v}_{e^*\varepsilon}(k) - \rho(k, \phi)\sigma^2/(1 - \phi^2)\}^2 \tag{2.17}$$

over a certain region of (ϕ, σ^2). Theoretically, one may choose any $K > 15$ (at least 3 equations for 2 parameters). For making ϕ to be identifiable from $\rho(k, \phi)$ and K should be moderate, we choose $K = 25$. ($\rho(25, 0) - \rho(25, 0.99) \approx 1$, but if $\rho(k, 0) - \rho(k, 0.99) \approx 0$, $k < 25$.)

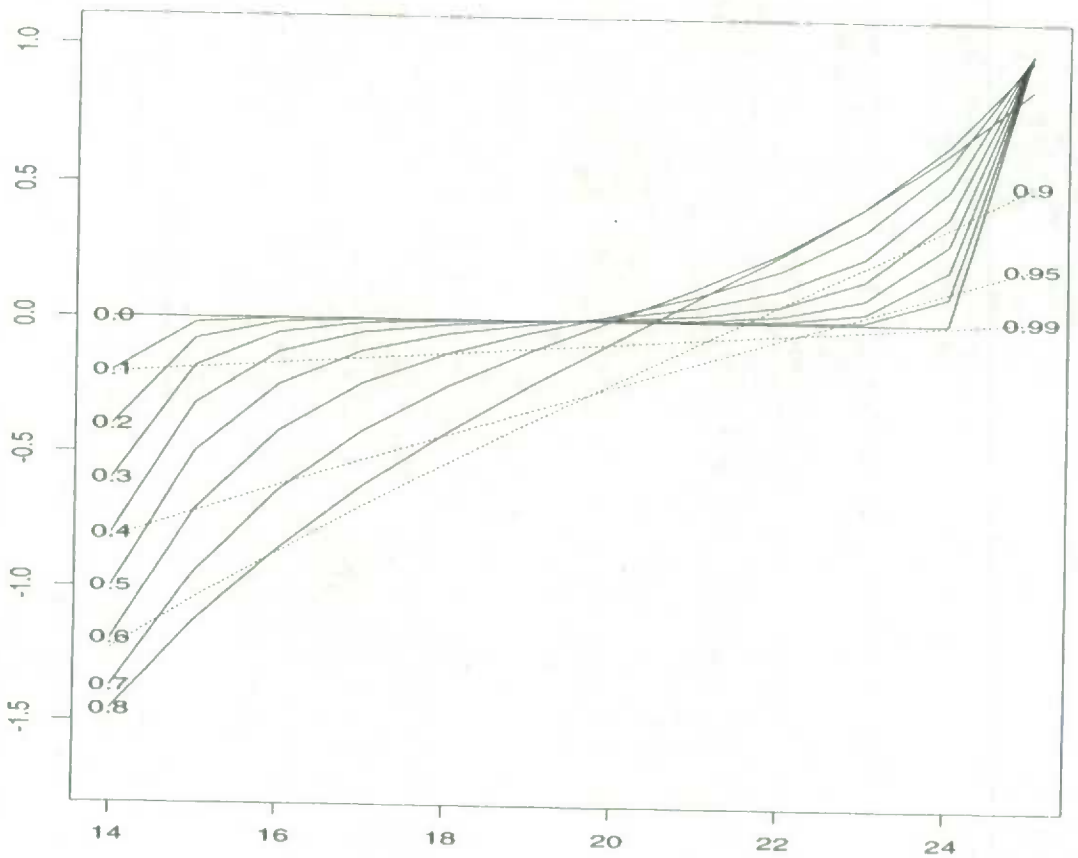

Figure 2.1. $\rho(k, \phi)$ versus k

The possible range for ϕ is from 0 to a number close to 1, say 0.99 because the autocorrelation of the survey error series is non-negative due to the overlapping of sample units. Suppose the estimate $\hat{v}_e(0)$ of $v_e(0)$ provided by survey experts has percentage error $\pm p\%$, i.e.

$$v_L = \hat{v}_e(0)(1 - p\%) \leq v_e(0) \leq \hat{v}_e(0)(1 + p\%) = v_U; \tag{2.18}$$

then from $\sigma^2 = v_e(0)(1 - \phi^2)$, the range of σ^2 will be

$$v_L(1 - \phi^2) \leq \sigma^2 \leq v_U(1 - \phi^2) \tag{2.19}$$

for the given ϕ, $0 \leq \phi \leq 0.99$. This yields the area for the 2-dimension minimization of (2.17). When $p = 0$, this area is degenerated to the line $\sigma^2/(1-\phi^2) = \hat{v}_e(0)$, $0 \leq \phi \leq 0.99$. A simple minimization method is the grid search procedure in the defined area which will be an easy task for modern computers.

2.5 Estimating $v_{e^*\varepsilon}(k)$ from data

The above discussion shows that for estimating (ϕ, σ^2), an estimate $\hat{v}_{e^*\varepsilon}(k)$ of $v_{e^*\varepsilon}(k)$ in (2.16) is required. To obtain such an estimate, we first suppose that $e^*(t) = e(t) - e(t-1) - e(t-12) + e(t-13)$ $(t = 14, 15, \cdots, n)$ are known quantities. Note that

$$E[e^*(12(T-1)+k)\varepsilon(T)] = E[e^*(k)\varepsilon(1)] = E[e^*(12(T+1)-(k-2))\varepsilon(T)]. \quad (2.20)$$

Since both $e^*(t)$ and $\varepsilon(t)$ are stationary, for $k = 14, \cdots, 24$, $v_{e^*\varepsilon}(k)$ can be estimated by

$$\hat{v}_{e^*\varepsilon}(k) = \frac{1}{2(N-1)} \left\{ \sum_{T=1}^{N-1} e^*(12(T-1)+k)\varepsilon(T) + \sum_{T=2}^{N} e^*(12(T+1)-(k-2))\varepsilon(T) \right\}. \quad (2.21)$$

For $k = 25, \cdots, 36$, the estimation formulae are the same except that the upper limit of the first summation is $N-2$, the lower limit of the second summation is 3 and the denominator is $2(N-2)$. Although $\hat{v}_{e^*\varepsilon}(k)$ can be obtained for higher values of k, but at most, we only use $k = 14, \cdots, 25$.

In practice, $\varepsilon(T)$ can be obtained from $y(t)$ and $z(T)$ by (2.5); but $e^*(t)$ is unknown. Since $\eta^*(t)$ and $e(t)$ are stationary, mean zero and mutually uncorrelated, as $N \to \infty$, we have

$$\frac{1}{2(N-1)} \left\{ \sum_{T=1}^{N-1} \eta^*(12(T-1)+k)\varepsilon(T) + \sum_{T=2}^{N} \eta^*(12(T+1)-(k-2))\varepsilon(T) \right\} \to 0, \quad (2.22)$$

for $k = 14, \cdots, 24$ (similar remark as above for $k > 24$). Hence, on the right hand side of (2.21), $e^*(t)$ can be replaced by $y^*(t)$. When N is not large enough, the condition of such a replacement has to be verified by simulation.

3. Estimation procedures in practical situations

In Subsection 3.1, we show that the 2-dimension simultaneous minimization of (2.17) [the optimal point is denoted by (ϕ_1, σ_1^2)] can be replaced by a two-step procedure of 1-dimension search using the relationship $\sigma^2 = v_e(0)(1-\phi^2)$. We show that both the simultaneous and the two-step procedures work well when N is large. However, when N is small, which is the case in practice, (ϕ_1, σ_1^2) becomes unacceptable. The reason is that the values of the target function for minimization are very close between very low ϕ and very high ϕ. When N is small, the fluctuation of $\hat{v}_{e^*\varepsilon}(k)$ is too large for distinguishing such a subtle difference and it leads to totally inappropriate estimates.

In Subsection 3.2, we transform the original target function for the minimization to another function which can clearly reflect the change of ϕ. However, when N is small, the fluctuation of its sample counterpart is still too large. Consequently, a procedure to trim the fluctuation (to adjust outliers) is introduced according to the assumption that we have an idea of the magnitude of ϕ being "low" or "high".

In Subsection 3.3, we introduce some auxiliary features (statistics) which can be used to test whether ϕ is 0, "low" or "high". It will be necessary if this information is not provided by survey experts. These features combined with the value of ϕ_1 are also used to obtain ϕ_2 in Subsection 3.4.

Subsection 3.4 introduces the estimate ϕ_2 of ϕ which is much better than ϕ_1. However, ϕ_2 is still considered as an intermediate estimate of ϕ but will be used to obtain a better estimate denoted by ϕ_3. Based on our simulation experience, we further revise ϕ_3 to get the final estimate denoted by ϕ_4 which will be discussed in Section 4.

After obtaining ϕ_4, we re-define a range for searching the optimal value of σ^2 by minimizing (2.17) with ϕ replaced by ϕ_4. Details on how to choose this range can be

found in Appendix C.

3.1 Problems with minimization when N is small

First, σ^2 is always fixed as 1. We generate $e(t)$ according to model (1.3) with $\phi = 0, 0.3, \ldots, 0.9, 0.95$. Then some $\eta(t)$ is added to $e(t)$ to get $y(t)$. Table 3.1 shows the mean and the standard deviation (denoted by Mean and SD respectively) of the estimated values of ϕ from 20 replications. A replication means using the same set of white noise $\xi(t) \sim \mathcal{N}(0,1)$ to obtain $e(t)$ defined in (1.3) by recursion with ϕ taking a value of $0, 0.3$ etc. For each replication, the estimates (ϕ_1, σ_1^2) are obtained by minimizing (2.17) simultaneously in the range defined by $0 \leq \phi \leq 0.99$ and σ^2 in (2.19) with $\hat{v}_e(0) = v_e(0)$ in (2.18). We do not display the result related to σ_1^2 since we take $p\% = 2\%$ so that $\sigma_1^2 \approx v_e(0)(1 - \phi_1^2)$. The rows of $N = 100$ and $N = 500$ are results for $N = 100$ and 500 with $\eta(t) = 0$ is added. The rows of 100(AR) and 500(AR) are the corresponding results when the added $\eta(t)$ follows model (2.10); where $\zeta(t)$ is a seasonal AR model as in (4.4) with $\phi_\eta = 0.8$, $\Phi_\eta = 0.6$, $\theta_\eta = \Theta_\eta = 0$, $\sigma_\eta^2 = 1$ (comparable with $\sigma^2 = 1$).

Table 3.1 shows that the results for $N = 500$ are excellent except for true $\phi = 0$ with $\eta(t)$ following the abovementioned AR model. In that case, only one replication produces $\phi_1 = 0.99$ which ruins the result. For $N = 100$, usually ϕ_1 is good, but in few replications for lower ϕ, ϕ_1 is far away from true ϕ which make SD too large. These results show that the estimation method in Section 2 works well when N is large. Unfortunately, in reality, for some national accounting problems, N could be as large as 20 but usually N is even smaller.

Table 3.1 Statistics of ϕ_1 for 20 replications

True ϕ		0	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$N = 100$	Mean	.189	.332	.390	.513	.601	.703	.790	.892	.949
	SD	.293	.281	.220	.212	.174	.126	.087	.039	.013
100(AR)	Mean	.194	.355	.425	.527	.631	.692	.779	.886	.948
	SD	.295	.249	.249	.205	.204	.194	.129	.061	.017
$N = 500$	Mean	.038	.286	.392	.497	.597	.694	.789	.898	.950
	SD	.065	.095	.069	.056	.050	.044	.037	.017	.006
500(AR)	Mean	.083	.301	.406	.508	.606	.702	.807	.897	.950
	SD	.224	.089	.070	.062	.057	.048	.033	.017	.008

Table 3.2 shows what could happen with N=10 where $\eta(t) = 0$. We did 100 replications with $p\% = 2\%$. 4 examples are cited in the table. #20 is one of few replications which provides reasonable results. Most other replications provide unacceptable results like those of #38, #65 and #98. The rows of Mean(1) and SD(1) give the mean and standard deviation of ϕ_1 of these 100 replications. Most of the values of SD(1) are even larger than the SD of the U[0,1], the uniform distribution on $[0, 1]$, which is equal to $(1/12)^{1/2} \simeq 0.29$. The rows of Mean(R) and SD(R) corresponding to the estimates from $R(k, \phi)$ will be explained in the next subsection.

Table 3.2 Some ϕ_1 and statistics of ϕ_1 and ϕ_R with N=10, 100 replications.

True ϕ	0	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
#20	0.38	0.44	0.48	0.54	0.64	0.85	0.84	0.86	0.91
#38	0.01	0.06	0.07	0.08	0.09	0.10	0.18	0.94	0.97
#65	0.10	0.43	0.53	0.61	0.66	0.69	0.69	0.65	0.51
#98	0.99	0.99	0.99	0.99	0.99	0.97	0.71	0.77	0.74
Mean(1)	0.41	0.49	0.55	0.57	0.62	0.68	0.73	0.80	0.86
SD(1)	0.42	0.38	0.36	0.35	0.33	0.29	0.26	0.21	0.21
Mean(R)	0.14	0.28	0.33	0.42	0.52	0.61	0.71	0.79	0.83
SD(R)	0.18	0.26	0.28	0.29	0.31	0.30	0.28	0.27	0.27

In fact, there is nothing to be surprised about the results in Table 3.2 if we look at contour plots of $S(\phi, \sigma^2)$ for some replications: most of the contours are almost parallel to the line $\sigma^2 = v_e(0)(1 - \phi^2)$, and the minimization of $S(\phi, \sigma^2)$ heavily depends on the values of p. This explains why the estimate ϕ_1 of ϕ is so volatile.

One may suspect that instead of using the simultaneous 2-dimension search, the 1-dimension search in two steps perhaps can improve the estimation. Suppose we know $v_e(0)$ exactly. Dividing both terms of (2.17) by $v_e(0) = \sigma^2/(1-\phi^2)$, ϕ_1 can be obtained by minimizing

$$S_\rho(\phi) = \sum_{k=14}^{25} \left\{\hat{\rho}(k) - \rho(k,\phi)\right\}^2, \ \hat{\rho}(k) = \hat{v}_{e^*\varepsilon}(k)/v_e(0), \ 0 \le \phi \le 0.99, \tag{3.1}$$

and $\sigma_1^2 = v_e(0)(1-\phi_1^2)$. [In practice, $v_e(0)$ is replaced by $\hat{v}_e(0)$; σ_1^2 is obtained by minimizing $S(\phi_1, \sigma^2)$ in (2.17) in the interval like (2.19). c.f. Appendix C] However, our simulation shows that the results are as bad as the 2-dimension search. The reason can be easily explained by Figure 2.1 which we have mentioned before: The lines for high ϕ and low ϕ heavily overlap. In particular, $\rho(k,0)$ and $\rho(k,0.99)$ are so close to each other for all k except $\rho(25,0) = 1$ and $\rho(25,0.99) = 0.013$. However, with $N = 10$, among our 20 simulation replications, $\hat{\rho}(25)$ can be as high as 5 and as low as -1 for $\phi = 0$. Thus, if $\hat{\rho}(25)$ is close to or less than 0, it is more likely to get $\phi_1 = 0.99$ rather than $\phi_1 = 0$.

3.2 Estimation strategies

As the first strategy to overcome the difficulties, we define $R(k,\phi)$ as follows: [See (2.14) and notice that $\sigma^2/(1-\phi^2) = v_e(0)$. In (3.2), the first expression is for theoretical discussion, the second expression is for obtaining some estimates of $R(k,\phi)$.]

$$R(k,\phi) = \rho(k,\phi)/G(\phi) = v_{e^*\varepsilon}(k,\phi,\sigma^2)/\{G(\phi)\sigma^2/(1-\phi^2)\} \tag{3.2}$$

where

$$G(\phi) = \{\frac{1}{11}\sum_{k=15}^{25}[\rho(k,\phi) - \rho(k-1,\phi)]^2\}^{1/2} \tag{3.3}$$

is a measure of the variation of $\rho(k,\phi)$ in some sense for a given ϕ. Some values of $G(\phi)$ are listed in Table 3.3. Note that $G(\phi) \le 0.302$ and decreases monotonically; it drops very fast when ϕ approaches 1. Therefore, the increase of the absolute value of $R(k,\phi)$ from the absolute value of $\rho(k,\phi)$ accelerates when ϕ increases, and extremely fast when ϕ approaches 1.

Table 3.3 Some values of $G(\phi)$.

ϕ	0	0.3	0.5	0.6	0.7	0.8	0.85	0.9	0.95	0.99
$G(\phi)$	0.302	0.258	0.247	0.240	0.233	0.217	0.198	0.159	0.094	0.020

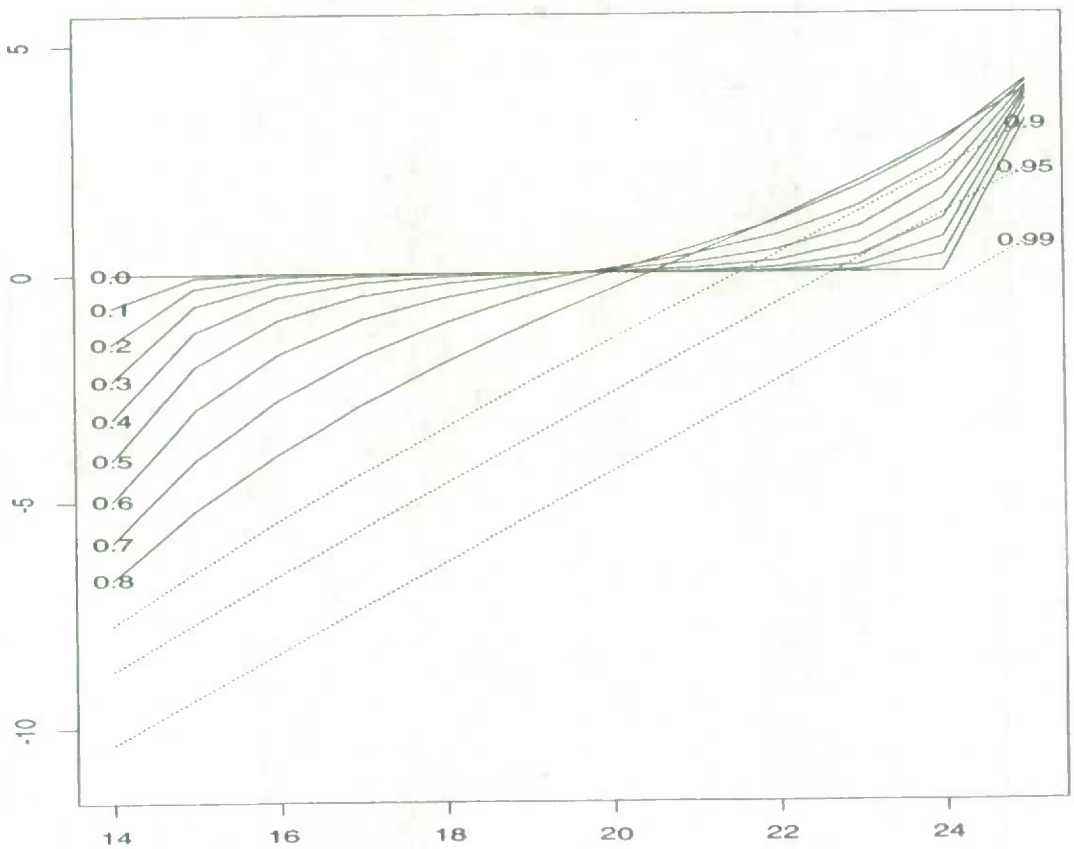

Figure 3.1. $R(k, \phi)$ versus k

Figure 3.1 is the plot of $R(k, \phi)$ versus k for $k = 14, \cdots, 25$. Within the range of $14 \leq k \leq 20$, any two $R(k, \phi)$ have almost no overlap and monotonic decreases. Thus instead of minimizing (3.1), we minimize (with $0 \leq \phi \leq 0.99$)

$$S_R(\phi) = \sum_{k=14}^{20} \{\hat{R}(k) - R(k, \phi)\}^2, \quad \hat{R}(k) = \hat{\rho}(k)/G(\phi_2), \tag{3.4}$$

where ϕ_2 is a reasonable estimate of ϕ (a better estimate than ϕ_1, see Subsection 3.4) and instead of impractical definition in (3.1), $\hat{\rho}(k)$ is defined as

$$\hat{\rho}(k) = \hat{v}_{e^*\varepsilon}(k)/\hat{v}_e(0). \tag{3.5}$$

Obviously, $R(k,\phi)$ is a much better target function than $\rho(k,\phi)$. However, our simulation examples shows that the variation of $\hat{R}(k)$ is still too large. The minimization of (3.4) is not sufficiently good to provide an acceptable estimate of ϕ.

Let us look at the two rows Mean(R) and SD(R) in Table 3.2 which are the mean and the standard deviation of the estimates, ϕ_R, obtained by minimizing $S_R(\phi)$ in (3.4) (in the expression of $\hat{R}(k)$, ϕ_2 temporarily takes the true value of ϕ). Obviously, ϕ_R is better than ϕ_1 for lower ϕ, especially those close to zero, but worse for higher ϕ. In fact, for higher ϕ, ϕ_R performs better than ϕ_1 when $\hat{\rho}(k) < 0$ for all k, $14 \leq k \leq 20$. But if $\hat{\rho}(k) > 0$ for some $k \leq 20$, these $\hat{R}(k)$ become "more positive" [the higher the ϕ_2, the smaller the $G(\phi_2)$] and $\hat{R}(k)$ may become closer to $R(k,0)$ rather than to $R(k,0.99)$. When $N = 10$, $\hat{\rho}(k) > 0$ may occur for any k, $14 \leq k \leq 20$ and any ϕ although it occurs more frequently for lower ϕ. In order to prevent ϕ_R from taking small values for higher ϕ, we may put $\hat{\rho}(k) = 0$ whenever $\hat{\rho}(k) \geq 0$ for $14 \leq k \leq 20$ in the minimization of (3.4). However, in this way, lower ϕ probably is aggressively over-estimated, the large bias and large SD still remain.

To overcome this difficulty, our second strategy is to determine a range, a subinterval in [0,0.99], in which the true ϕ falls. Sometimes, survey experts may provide information of whether a survey error series, $e(t)$, is uncorrelated, weakly correlated or strongly correlated. Sometimes, even a range of ϕ may be provided. For examples (unfortunately, it is not a good example for the benchmarking model in this paper) in the Labor Force Survey in Canada, ϕ most likely falls within [0.35,0.6] for unemployment series and falls within [0.6,0.85] for employment series. However, if such information is unavailable, the method proposed in the next subsection provides us with a tool to make such a judgment. Since there is no need to model an uncorrelated series, the range of ϕ is divided into two different parts, namely "low ϕ" and "high ϕ" which represent, say, $0 < \phi \leq 0.7$ and $0.7 \leq \phi \leq 0.99$ respectively. On one hand, if $\phi > 0$ but close to 0, say $\phi < 0.4$, we may simply assign $\phi = 0$ as it almost makes no difference when the model is used for signal extraction or benchmarking. On the

other hand, if ϕ is very close to 1, the survey error behaves as a random walk which is not supposed to happen. Thus we may assume $\phi = 0.95$ is the upper bound of ϕ. Therefore we may assume "low ϕ" and "high ϕ" as $0.4 \leq \phi \leq 0.7$ and $0.7 \leq \phi \leq 0.95$ respectively. These two special definitions are just for convenience of statement in the rest of this paper to demonstrate how we carry out the procedures and what results we may obtain. The parameters in the estimation process should be changed accordingly if the assumed range of the true ϕ is different from these two intervals.

The third strategy is to confine $\hat{R}(t)$ according to the assumed range of ϕ. For a fixed k in [14, 20], $R(k, \phi)$ is a decreasing function in ϕ and is negative for all $\phi > 0$. Assume the true ϕ falls in the interval $[\phi'_{min}, \phi'_{max}]$. [If the interval is like the abovementioned "high ϕ", we treat it slightly differently, see (4.1).] Then we confine $\hat{R}(k)$ in the range targeting $[R(k, \phi'_{max}), R(k, \phi'_{min})]$ by trimming $\hat{\rho}(k)$ to $\rho^*(k)$ as follows:

$$\rho^*(k) = \begin{cases} \hat{\rho}(k), \text{ if } \rho(k, \phi_{max}) < \hat{\rho}(k) < \rho(k, \phi_{min}); \\ \rho(k, \phi_{max}), \text{ if } \hat{\rho}(k) \leq \rho(k, \phi_{max}); \\ \rho(k, \phi_{min}), \text{ if } \hat{\rho}(k) \geq \rho(k, \phi_{min}), \end{cases} \tag{3.6}$$

where ϕ_{min} (or ϕ_{max}) is around ϕ'_{min} (or ϕ'_{max}) and will be determined by simulation (see Appendix D). Thus instead of minimizing (3.4), we minimize ($0 \leq \phi \leq 0.99$)

$$S_R^*(\phi) = \sum_{k=14}^{20} \{R^*(k) - R(k, \phi)\}^2, \quad R^*(k) = \rho^*(k)/G(\phi_2). \tag{3.7}$$

3.3 Auxiliary features, test for $\phi = 0$, "low ϕ" or "high ϕ"

Due to the large volatility of ϕ_1, we cannot simply let $\phi_2 = \phi_1$ in $G(\phi_2)$ in (3.7). In fact, ϕ_1 cannot be used as a guideline to judge whether ϕ is "low" or "high". In this subsection, some auxiliary features are employed to judge whether ϕ is zero, "low" or "high", in the cases when this information is not provided by survey experts.

First note that, as a function of k, $\hat{v}_{\varepsilon^*\varepsilon}(k)$ given in (2.21) fluctuates in similar magnitudes for all ϕ. But $\hat{\rho}(k)$ in (3.1) does not behave in the same manner: its fluctuation is usually much larger when ϕ is low than when ϕ is high. The reason is

simple: For a fixed σ^2, $v_e(0) = \sigma^2/(1-\phi^2)$ tends to infinity as ϕ tends to 1 (note that $\hat{\rho}(k)$ defined in (3.1) should not depend on σ^2). We define

$$SC = \sum_{k=15}^{20} \{\hat{\rho}(k) - \hat{\rho}(k-1)\}^2 \tag{3.8}$$

to measure this change, where SC stands for "squared change". Alternatively, we could take 25 as the upper limit of the summation or employ the absolute change rather than the squared change. As indicated in Table 3.4, the Mean and the SD of SC almost remain unchanged for $\phi < 0.7$ and decline faster and faster for $\phi \geq 0.7$. For $\phi = 0.99$, the Mean and SD of SC are 0.66 and 0.74 respectively. Therefore SC may serve as a tool to differentiate "high ϕ" from "low ϕ". However due to large SD of SC, the results are only reliable for very large ϕ, for instance, $\phi \geq 0.9$.

Table 3.4 Statistics for SC, LV and A (10,000 replications, $\eta(t) = 0$)

	True ϕ	0	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
SC	Mean	43.9	48.1	48.2	47.4	45.1	40.4	31.6	16.2	6.7
	SD	46.0	48.1	47.7	46.2	43.3	38.0	29.0	14.7	6.2
LV	Mean	-0.64	-1.33	-1.74	-2.29	-3.02	-3.93	-4.84	-4.79	-3.48
	SD	1.19	1.71	1.94	2.20	2.51	2.85	3.09	2.75	1.96
A	Mean	1.68	1.77	1.83	1.91	2.00	2.13	2.28	2.43	2.48
	SD	0.22	0.27	0.30	0.30	0.31	0.31	0.30	0.27	0.28

Second, recall that (see Figure 2.1) for $14 \leq k \leq 20$, when ϕ increases from 0, $\rho(k, \phi)$ decreases to its lowest level around $\phi = 0.8$ or 0.9 and then rebounces. Thus we expect that the average level of $\hat{\rho}(k)$ behaves in a similar fashion. Define this average as (LV stands for "level")

$$LV = \sum_{k=15}^{20} w(k)\hat{\rho}(k), \tag{3.9}$$

where $w(k)$ are some weights to be chosen. According to Figure 2.1, the left hand side ($k = 14, 15$, etc) is more important than the right hand side ($k = 20, 19$, etc). However, positive values of $\hat{\rho}(14)$ and $\hat{\rho}(15)$ lead to a more serious error for large ϕ. Hence we suggest to use conditional weights such as (3.10). The criterion for choosing $w(k)$ is to make the range of mean of LV as large as possible when ϕ goes from 0

to 0.8 with possibly the smallest SD. After extensive trials, we decide to adopt the following weights which give the smallest SD among several choices:

$$\begin{cases} w(14) = 1/3 & \text{if } \hat{\rho}(14) > 0, & \text{or} & 1 \text{ otherwise;} \\ w(15) = 2/3 & \text{if } \hat{\rho}(15) > 0, & \text{or} & 1 \text{ otherwise;} \\ w(t) = 1 & 16 \leq t \leq 18; & w(19) = 2/3; & w(20) = 1/3. \end{cases} \tag{3.10}$$

With weights given in (3.10), the simulation results of LV are provided in Table 3.4. When ϕ increases from 0, the Mean of LV decreases until ϕ is somewhere between 0.8 and 0.9 and then increases, producing similar mean for $\phi = 0.7$ and for $\phi = 0.95$. This is a serious drawback of LV. Also, the SD of LV tends to be slightly larger.

Both SC and LV are not ideal to reflect the change of ϕ in the entire range of [0,0.99]. A statistic which performs better than both SC and LV, denoted by A (standing for "angle" or "arc cos") is defined from SC and LV as

$$A = \text{arc cos}(\sqrt{SC}/LV). \tag{3.11}$$

Since $\sqrt{SC} > 0$, A takes values in $(0, \pi)$ and mostly falls in $(\pi/2, \pi)$ as LV usually takes negative values if ϕ is far away from zero. From Table 3.4, we see that the mean of A increases monotonically with ϕ in the range of $[0, 0.95]$ with relatively small SD. The Mean for 0.5 is totally outside the intervals of Mean $\pm SD$ for $\phi = 0.8$ and vice versa. Table A1 in Appendix A is the "histograms" of A for some typical values of ϕ in the case $\eta(t) = 0$ which gives the numbers of replications among 10,000 that A falls in the intervals listed in the leftmost column. The length of each interval is $\pi/40 \approx 0.08$. The last two rows in Table A1 are the same as the last two rows in Table 3.4.

Based on Appendix A, we construct Table 3.5 where A_C is some critical values. The numbers under ϕ values are the sizes or powers associated with the corresponding ϕ when we set the null hypothesis as "H_0: ϕ is high". Suppose 0.7 is considered as the lower bound of "high ϕ". We reject H_0 when $A < A_C$. Then the numbers in the columns of $\phi \geq 0.7$ are sizes and those in the columns of $\phi < 0.7$ are powers. For keeping small sizes, $A_C = 1.96$ is preferred. The change of these sizes and powers

depends on the model of $\eta(t)$ when it is non-zero, but the change is usually not very large if the model parameters are reasonable.

If the hypothesis is "$H_0 : \phi$ is low", then H_0 will be rejected when $A \geq A_C$. Denote the numbers in Table 3.5 by α, then $1-\alpha$ is the size for "low ϕ" and $1-\alpha$ is the power for "high ϕ". For controlling the type I error, $A_C = 2.28$ is preferred. For testing "$H_0 : \phi = 0$" only, we may take $A_C = 1.96$, then the size is $1 - 0.91 = 0.09$. The powers for $\phi = 0.5$ and 0.9, for examples, are $1 - 0.58 = 0.42$ and $1 - 0.06 = 0.94$ respectively. In fact, the information of $\phi = 0$ or $\phi > 0$ usually can be shown according to whether there is overlapping or not of sample units or panels in the survey design. In most situations, only testing "high ϕ" versus "low ϕ" is in need.

Table 3.5 Critical values of A and associated sizes and powers

	ϕ								
A_C	0	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
1.96	0.91	0.78	0.69	0.58	0.44	0.29	0.14	0.06	0.05
2.12	0.98	0.91	0.85	0.77	0.64	0.47	0.28	0.13	0.10
2.28	0.99	0.97	0.94	0.89	0.81	0.67	0.47	0.26	0.20

3.4 A reasonable estimate ϕ_2 of ϕ

This estimate is obtained by referring to "low ϕ" or "high ϕ", but we also count possible mis-identification. According to SD(1) in Table 3.2, ϕ_1 is volatile and it is much worse for "low ϕ" than for "high ϕ". For a "low ϕ", say $\phi = 0.5$, ϕ_1 spreads all over $[0, 0.99]$ and takes extremely small or large values with large probability. However the performance of A is very good no matter ϕ is high or low. Note that A concentrates around its mean corresponding to the true ϕ, and the mean is almost identical with the mode (see Appendix A). In fact, using A to estimate ϕ is quite reliable.

For $\phi \in (0, 0.8)$, $G(\phi)$ decreases slowly as ϕ increases. If the estimate, ϕ_2, of a very low ϕ, say $\phi = 0.3$, is not too high, say $\phi_2 = 0.8$, or vice versa, the impact on $\hat{R}(k) = \hat{\rho}(k)/G(\phi_2)$ [see (3.4)] is not very serious. Hence, for "low ϕ", we simply set

$$\phi_2 = \begin{cases} 0.2, & \text{if } A < M_{0.3}(A) = 1.77, \\ j, & \text{if } M_j(A) \leq A < M_{j+0.1}(A), \quad j = 0.3, \cdots, 0.7; \\ 0.8, & \text{if } A \geq M_{0.8}(A) = 2.28. \end{cases} \tag{3.12}$$

where $M_j(A)$ is the Mean of A in Table A1 when $\phi = j$. From Table A1, we see that the probability of having serious mistake is small: The probability is about $1.5\% \sim 4\%$ of getting $\phi_2 \geq 0.8$ for $\phi \leq 0.3$ or $\phi_2 \leq 0.3$ for $\phi \geq 0.8$. The simulation in the next section shows that ϕ_2 is much better than ϕ_1.

The situation for "high ϕ" is somewhat different. On one hand, $G(\phi)$ drops rapidly when $\phi \geq 0.8$. For example, getting $\phi_2 = 0.9$ for $\phi = 0.7$ does have some impact on $\hat{R}(t)$. On the other hand, from Table 3.4, for $\phi > 0.7$, the change of the mean of SC is much more significant than the change of the mean of A. So the information carried by SC is now useful. Furthermore, for very high ϕ (≥ 0.9), the SD of ϕ_1 becomes smaller; and in fact, with only few replications ϕ_1 being too low. [That makes SD(1) = 0.21 in Table 3.2. Otherwise, SD(1) will be much smaller than 0.21.] So ϕ_1 is useful for the "high ϕ" situation. Appendix B describes the steps of how to use A and SC as well as ϕ_1 to obtain ϕ_2.

4. Simulation

In this section, we focus only on the estimation of ϕ. The estimation of σ^2 is demonstrated in Appendix C. All simulation results are obtained from using N=10 with $\sigma^2 = 1$ for model (1.3).

4.1 Simulation results: the case where $\eta(t) = 0$

In this subsection, we set $\eta(t) = 0$ in (1.1). Thus $z(T) = 0$ as we only consider

binding benchmarks. The purpose of this subsection is to demonstrate how the procedures work without the effect of $\eta(t)$.

We regard $0.5 \sim 0.6$ as the core range of "low ϕ", 0.4 and 0.7 are lower and upper bounds of "low ϕ", but the simulation will also be carried out for the extended range of true ϕ at values 0.3 and 0.8. For "high ϕ", we regard $0.8 \sim 0.95$ as the core range, 0.7 is the lower bound, but we will extend it to 0.6, even to 0.5. The results in the extended ranges may shed some light on the behavior of the estimate if ϕ is mis-identified around 0.7.

The first step is using 2-dimensional search to get (ϕ_1, σ_1^2) (or using 1-dimensional search to get ϕ_1 only). The second step is using A, SC and ϕ_1 to obtain a reasonable estimate ϕ_2. The third step is to trim $\hat{\rho}(k) = \hat{v}_{e^*\varepsilon}(k)/\hat{v}_e(0)$ to $\rho^*(k)$ according to (3.6) and obtain ϕ_3 by minimizing (3.7). For "high ϕ", (3.6) should be revised because it is hard to see which ϕ gives the lowest $\rho(k, \phi)$ with $14 \leq k \leq 20$. Figure 2.1 shows it could be $\phi = 0.8$, 0.9 or 0.95 (or a value between them). It seems that ϕ_{max} should vary with k. Usually, the best results are obtained for those ϕ around the middle of (ϕ_{min}, ϕ_{max}). Since $0.8 \sim 0.95$ are the core range of "high ϕ", simply trimming $\hat{\rho}(k)$ by (3.6) may not lead to good results for ϕ in this range. Thus the following formula is used for "high ϕ":

$$\rho^*(k) = \begin{cases} \hat{\rho}(k), & \text{if } \alpha\rho(k, \phi_{max}) < \hat{\rho}(k) < \rho(\phi_{min}); \\ \alpha\rho(k, \phi_{max}), & \text{if } \hat{\rho}(k) \leq \alpha\rho(k, \phi_{max}); \\ \rho(k, \phi_{min}), & \text{if } \hat{\rho}(k) \geq \rho(k, \phi_{min}). \end{cases} \tag{4.1}$$

where, $\alpha \geq 1$ is also chosen by simulation (see Appendix D).

The fourth step which is totally based on our experience. Suppose we know the true value of $v_e(0) = \sigma^2/(1-\phi^2)$ (ϕ and σ^2 are unknown), then we may put $\hat{\rho}(k) = \hat{v}_{e^*\varepsilon}(k)/v_e(0)$. However, our simulation shows that, instead of $v_e(0)$, using $\sigma^2/(1-\phi_2^2) = v_e(0)(1-\phi^2)/(1-\phi_2^2)$ as denominator to calculate $\hat{\rho}(k)$ [see (3.1)] provides better results. Since $v_e(0)$ is estimated by $\hat{v}_e(0)$ and ϕ_3 is considered to be

the best estimate of ϕ, thus practically, we may put

$$\hat{\rho}(k) = \hat{v}_{e^*\varepsilon}(k)/\{\hat{v}_e(0)(1-\phi_3^2)/(1-\phi_2^2)\}. \tag{4.2}$$

After obtaining ϕ_3, we re-define $\hat{\rho}(k)$ as (4.2), then get $\rho^*(k)$ from (3.6) for "low ϕ" or from (4.1) for "high ϕ", and redo the minimization (3.7) to obtain an optimal value denoted by ϕ_4 which is the final estimate of ϕ.

In all steps, the estimate $\hat{v}_e(0)$ and p play important roles:

1. For obtaining (ϕ_1, σ_1^2), $\hat{v}_e(0)$ and $p\%$ define the minimization area.
2. $\hat{v}_e(0)$ is used to obtain $\hat{\rho}(t)$ from $\hat{v}_{e^*\varepsilon}(t)$.
3. $\hat{v}_e(0)$ and $p\%$ determine the minimization range for obtaining σ_4^2 (see Appendix C), the final estimate of σ^2.

Usually, $\hat{v}_e(0)$ is published by statistical agencies but the percentage error of $\hat{v}_e(0)$, i.e. $|\hat{v}_e(0) - v_e(0)|/v_e(0) \leq p\%$, is not given. Fortunately, p has almost no effect on ϕ_3 and ϕ_4 if $p \leq 50$. Its effect on σ_4^2 (or σ_3^2) is very mild, although the result is better for smaller p.

In our simulation, we choose $p = 2,\ 20$ and 50 which correspond to $\hat{v}_e(0)$ being an almost precise, quite good and rough estimate of $v_e(0)$. We generate $\hat{v}_e(0)$ according to the distribution $\mathcal{N}(v_e(0), (v_e(0)p/200)^2)$, i.e. the standard deviation of the percentage error is $(p/2)\%$. Thus with probability more than 0.95, $|\hat{v}_e(0) - v_e(0)|/v_e(0) \leq p\%$. $p = 0$ means $\hat{v}_e(0) = v_e(0)$.

Table 4.1 and 4.2 show the statistics of the estimates of ϕ in all steps based on 10,000 replications. Let $\phi_j^{(i)}$ be the value of ith replication of estimate ϕ_j, $j = 1,\ 2,\ 3,\ 4$; the RMSE (root mean squares error) of ϕ_j is defined as

$$\text{RMSE} = \{\sum_{i=1}^{10000} (\phi_j^{(i)} - \phi)^2/10000\}^{1/2} = \{(\text{Mean} - \phi)^2 + SD^2\}^{1/2}. \tag{4.3}$$

The conditions for these results are: $p = 20$ (the results are almost the same if we choose $p = 2$ or $p = 50$); $\hat{v}_e(0) \sim \mathcal{N}(v_e(0), (v_e(0)p/200)^2)$; $\phi_{min} = 0.4$ and $\phi_{max} = 0.7$ in (3.6) for "low ϕ", or $\phi_{min} = 0.5$ and $\phi_{max} = 0.9$ with $\alpha = 1.5$ in (4.1) for "high ϕ". In Appendix D, we show how the choice of these parameters affects these statistics and carry on more discussions.

From these tables, we see that both RMSE and SD of ϕ_1 are unacceptable. Those of ϕ_2 are much better but their RMSE and SD are still large. The improvement of ϕ_3, and especially ϕ_4 is significant. In the core ranges of "low ϕ" and "high ϕ", the RMSE of ϕ_3 and ϕ_4 can be much less than 0.1 with very small bias. Such small RMSE can be attained by ϕ_1 only for N as large as 500 (even $N = 100$ is not enough). In these ranges, ϕ_4 and ϕ_3 are quite the same for "low ϕ" (ϕ_4 is better around the upper bound), but ϕ_4 is significantly better than ϕ_3 for "high ϕ" in its core range. For more discussions, see Appendix E.

Table 4.1 statistics of ϕ estimates, "low ϕ", $\eta(t) = 0$

	ϕ	0.3	0.4	0.5	0.6	0.7	0.8
Mean	ϕ_1	0.486	0.520	0.561	0.620	0.676	0.742
	ϕ_2	0.350	0.392	0.447	0.515	0.598	0.680
	ϕ_3	0.521	0.532	0.545	0.561	0.581	0.602
	ϕ_4	0.527	0.539	0.554	0.573	0.596	0.624
RMSE	ϕ_1	0.432	0.391	0.361	0.329	0.295	0.259
	ϕ_2	0.189	0.199	0.220	0.234	0.230	0.211
	ϕ_3	0.229	0.146	0.080	0.079	0.138	0.209
	ϕ_4	0.235	0.152	0.083	0.069	0.121	0.185
SD	ϕ_1	0.390	0.373	0.356	0.329	0.294	0.252
	ϕ_2	0.182	0.199	0.213	0.219	0.206	0.173
	ϕ_3	0.061	0.064	0.066	0.069	0.071	0.069
	ϕ_4	0.059	0.061	0.063	0.063	0.063	0.059

Table 4.2 statistics of ϕ estimates for "high ϕ", $\eta(t) = 0$

	ϕ	0.5	0.6	0.7	0.8	0.9	0.95
Mean	ϕ_1	0.561	0.620	0.676	0.742	0.815	0.856
	ϕ_2	0.635	0.654	0.682	0.723	0.815	0.891
	ϕ_3	0.704	0.721	0.743	0.774	0.825	0.871
	ϕ_4	0.725	0.748	0.777	0.819	0.882	0.932
RMSE	ϕ_1	0.361	0.329	0.295	0.259	0.227	0.238
	ϕ_2	0.165	0.117	0.116	0.146	0.147	0.110
	ϕ_3	0.218	0.146	0.096	0.098	0.133	0.147
	ϕ_4	0.238	0.167	0.111	0.084	0.087	0.078
SD	ϕ_1	0.356	0.329	0.294	0.252	0.211	0.219
	ϕ_2	0.096	0.104	0.115	0.124	0.120	0.093
	ϕ_3	0.077	0.082	0.086	0.095	0.110	0.124
	ϕ_4	0.077	0.079	0.080	0.082	0.085	0.076

4.2. Simulation results: the case where $\eta(t)$ is nonstationary

The proposed survey error modelling procedure should work when a nonstationary series $\eta(t)$ is added to $e(t)$ as (1.1). Otherwise it is useless. An advantage of this procedure is that the model specification of $\eta(t)$, like (4.4) or (4.5) below, is not involved in the estimation process. However the specification does affect the error of the estimates of ϕ and σ^2. In order to investigate the impact of variables, we set different models and parameter specifications for $\eta(t)$.

First, we assume that $\eta(t)$ follow a TS model as (2.10) with $\eta^*(t) = \nabla\nabla_{12}\eta(t) = \nabla\nabla_{12}\zeta(t)$ and $\zeta(t)$ follow a seasonal ARMA model as

$$(1 - \phi_\eta B)(1 - \Phi_\eta B^{12})\zeta(t) = (1 - \theta_\eta B)(1 - \Theta_\eta B^{12})a_\eta(t) \tag{4.4}$$

where $a_\eta(t)$ is a white noise of mean zero and variance σ_η^2. Hence, $\eta^*(t)$ is over-differenced in two times. Second, we assume that $\eta(t)$ follow the DS model as (2.9) and

$$(1 - \phi_\eta B)(1 - \Phi_\eta B^{12})\eta^*(t) = (1 - \theta_\eta B)(1 - \Theta_\eta B^{12})a_\eta(t). \tag{4.5}$$

Then, $\eta^*(t)$ is not over-differenced. Also, we consider two sets of coefficients:

$$\phi_\eta = 0.8,\ \Phi_\eta = 0.6,\ \theta_\eta = \Theta_\eta = 0, \quad \text{(seasonal AR model)} \tag{4.6}$$

or

$$\phi_\eta = \Phi_\eta = 0,\ \theta_\eta = 0.8,\ \Theta_\eta = 0.6 \quad \text{(seasonal MA model)} \tag{4.7}$$

The combinations of (4.4) or (4.5) with (4.6) or (4.7) represent many situations in reality.

Table 4.3 RMSE and other statistics of ϕ_4, "low ϕ"

Conditions		Mean Diff.	0.4	0.5	0.6	0.7	0.8	SD Range
$\sigma_\eta = 0$	$\eta(t) = 0$	0.057	0.152	0.083	0.069	0.121	0.185	0.059-0.063
$\sigma_\eta = 1/3$	TS + AR	0.055	0.152	0.082	0.069	0.124	0.188	0.059-0.063
	TS + MA	0.048	0.151	0.078	0.067	0.128	0.200	0.057-0.059
	DS + AR	0.050	0.165	0.101	0.088	0.136	0.204	0.079-0.085
	DS + MA	0.055	0.153	0.084	0.071	0.123	0.187	0.059-0.061
$\sigma_\eta = 1$	TS + AR	0.045	0.154	0.081	0.072	0.132	0.203	0.061-0.067
	TS + MA	0.022	0.151	0.071	0.069	0.146	0.233	0.052-0.053
	DS + AR	0.028	0.189	0.130	0.123	0.171	0.244	0.113-0.123
	DS + MA	0.043	0.153	0.080	0.071	0.131	0.202	0.058-0.062
$\sigma_\eta = 3$	TS + AR	0.022	0.161	0.087	0.080	0.149	0.232	0.068-0.071
	TS + MA	0.009	0.150	0.066	0.073	0.158	0.254	0.049-0.050
	DS + AR	0.010	0.206	0.148	0.141	0.192	0.269	0.135-0.137
	DS + MA	0.023	0.154	0.076	0.071	0.146	0.232	0.057-0.059

Table 4.4 RMSE and other statistics of ϕ_4, "high ϕ"

Conditions		Mean Diff.	0.6	0.7	0.8	0.9	0.95	SD Range
$\sigma_\eta = 0$	$\eta(t) = 0$	0.166	0.182	0.129	0.101	0.087	0.061	0.060-0.087
$\sigma_\eta = 1/3$	TS + AR	0.160	0.181	0.127	0.097	0.089	0.067	0.067-0.092
	TS + MA	0.136	0.171	0.104	0.073	0.096	0.093	0.073-0.084
	DS + AR	0.114	0.196	0.143	0.118	0.131	0.139	0.113-0.130
	DS + MA	0.163	0.183	0.124	0.094	0.089	0.066	0.066-0.096
$\sigma_\eta = 1$	TS + AR	0.125	0.174	0.113	0.084	0.101	0.101	0.084-0.088
	TS + MA	0.047	0.159	0.079	0.064	0.128	0.160	0.054-0.066
	DS + AR	0.061	0.206	0.159	0.156	0.195	0.220	0.148-0.176
	DS + MA	0.122	0.169	0.104	0.076	0.099	0.102	0.074-0.083
$\sigma_\eta = 3$	TS + AR	0.038	0.163	0.092	0.080	0.142	0.177	0.072-0.080
	TS + MA	0.010	0.159	0.073	0.068	0.149	0.193	0.050-0.054
	DS + AR	0.020	0.201	0.161	0.171	0.227	0.262	0.158-0.169
	DS + MA	0.030	0.161	0.084	0.067	0.135	0.174	0.058-0.065

Table 4.3 and 4.4 listed the RMSE and other statistics based on 10,000 replications for all these combinations. For each replication, we generate Gaussian white noise $\xi(t)$ and $a(t)$ independently with mean 0 and $Var(\xi(t)) = Var(a(t)) = \sigma^2 = 1$. Using (1.3), $e(t)$ can be obtained from $\xi(t)$ recursively. Putting $a_\eta(t) = \sigma_\eta a(t)$ (where $\sigma_\eta = 1/3$, 1, or 3), we have $Var(a_\eta(t)) = \sigma_\eta^2$. For TS model, we may simply let $\eta(t) = \zeta(t)$ [the linear trend and the seasonality in (2.10) do not affect the procedure] which can be obtained from $a_\eta(t)$ using (4.4); for DS model, we obtain $\eta^*(t)$ from $a_\eta(t)$ using (4.5), then integrate $\eta^*(t)$ to $\eta(t)$. Finally, monthly survey data, $y(t) = \eta(t) + e(t)$ are created. The annual binding benchmarks, $z(T)$, are obtained from $\eta(t)$ using (1.2) [put $\psi(T) = 0$]. The ratio $\sigma_\eta^2/\sigma^2 = 1/9$, 1 and 9 may represent situations of low, medium and high signal-to-noise ratio respectively. Here "signal" means the stochastic variation in $\eta(t)$. It was pointed out by some authors, e.g., Trabelsi and Hillmer (1990), Chen, Cholette and Dagum (1997) that the lower the signal-to-noise ratio, the more the prediction error for $\eta(t)$ can be reduced by benchmarking via signal extraction. In very high signal-to-noise ratio situation, benchmarking via signal extraction and via regression gives almost the same results.

In either Table 4.3 or Table 4.4, the numbers in the first row are the true values

of ϕ. Assume 0.7 to be the bound of "low ϕ" and "high ϕ" and extend it by 0.1 to 0.8 for "low ϕ" and to 0.6 for "high ϕ. The second row ($\sigma_\eta = 0$) shows statistics when $\eta(t) = 0$ which provides a reference for other rows of non-zero $\eta(t)$ in different situations. TS and DS correspond to (4.4) and (4.5), +AR and +MA represent the choice of coefficients following (4.6) and (4.7) respectively.

The numbers in the columns under the true values of ϕ are RMSE of ϕ_4 for the given situations (σ_η, TS or DS, AR or MA). The numbers in the column of "SD Range" give the range of standard deviations of ϕ_4 for the given situation with ϕ taking all the values listed in the first row. The numbers in the column of "Mean Diff." are the differences between means of ϕ_4 when $\phi = 0.7$ and when $\phi = 0.4$ for "low ϕ", and $\phi = 0.95$ and 0.7 for "high ϕ". A smaller "Mean Diff." means, for all true ϕ between 0.4 and 0.7, or 0.7 and 0.95, ϕ_4 is more concentrated around somewhere (having minimum bias) in the corresponding region. For further discussions, see Appendix F.

5. Concluding remarks

In this paper, we proposed a new method to model monthly survey error, $e(t)$, when annual benchmarks are available. We assumed the survey error follows an $AR(1)$ model as in (1.3) and its variance estimate, $\hat{v}_e(0)$, is published by statistical agencies. The assumption for the variable of interest $\eta(t)$ is very general: $\eta^*(t) = \nabla\nabla_{12}\eta(t)$ is stationary (allowed to be over-differenced). Any specification of $\eta(t)$ or $\eta^*(t)$ does not affect our modelling procedure.

From monthly survey data $y(t) = \eta(t) + e(t)$ and annual benchmarks $z(T) = \Sigma_{t\in T}\eta(t)$, assuming $\eta^*(t)$ and $e(t)$ are mutually uncorrelated, we derived the theoretical expression [given by (2.14)] of $v_{y^*\varepsilon}(k) = v_{e^*\varepsilon}(k)$, the covariance of $y^*(t) = \nabla\nabla_{12}y(t)$ and $\varepsilon(T) = \Sigma_{t\in T}e(t) = \Sigma_{t\in T}y(t) - z(T)$. Since its sample covariance $\hat{v}_{y^*\varepsilon}(k)$ can be obtained by (2.21) (with e^* replaced by y^* all through in the formulae), we use a

non-linear regression model to obtain the least squares estimate (ϕ_1, σ_1^2) of (ϕ, σ^2) simultaneously by the minimization of (2.17). This 2-dimension minimization provides similar result as ϕ_1 obtained from 1-dimension minimization of (3.1) [with $v_e(0)$ replaced by $\hat{v}_e(0)$]. Our simulation shows that both procedures only work well for sample size as large as for hundreds years which is impractical. In real situations, only few annual benchmarks, at most about 20, are available.

The failure of the small sample performance of the estimator is due to two major reasons:

1. The heavy overlap or close values of the target function $\rho(k, \phi)$ between high ϕ and low ϕ.

2. The variation of sample counterpart, $\hat{\rho}(k)$, of $\rho(k, \phi)$ is too large such that the level and the pattern of $\rho(k, \phi)$ is submerged in it.

Three strategies are used to overcome the abovementioned difficulties:

1. Transform the target function $\rho(k, \phi)$ to $R(k, \phi) = \rho(k, \phi)/G(\phi)$. For $k = 14, \cdots, 20$, $R(k, \phi)$ monotonically decreases with increasing ϕ, $0 \leq \phi \leq 0.99$.

2. Get some information about possible subinterval in which the true ϕ falls. Whenever this information is not provided by survey experts, auxiliary features of $\hat{\rho}(k)$ such as A which can be used to test whether $\phi = 0$, "low ϕ" or "high ϕ". With the help of these features as well as ϕ_1, a preliminary reasonable estimate ϕ_2 is obtained (see Subsection 3.4).

3. Trim $\hat{\rho}(k)$ to $\rho^*(k)$ as (3.6) or (4.1) accordingly and use $R^*(k) = \rho^*(k)/G(\phi_2)$ for minimization [see (3.7)] to obtain ϕ_3.

Our experience shows that a technical adjustment [see (4.2)] leads to further improvement in which this final estimate of ϕ is denoted by ϕ_4. Then σ_4^2, an estimate of σ^2 based on ϕ_4, is obtained by minimizing (2.17) with ϕ replaced by ϕ_4. See Appendix

C for details.

The performance of the proposed procedure is first tested by assuming $\eta(t) = 0$ with different choice of parameters (some threshold and reference values in the procedures). Simulation shows that the bias of ϕ_4 is very small when the true ϕ falls in the core range of "low ϕ" or "high ϕ". As the SD (standard deviation) of ϕ_4 is always small, so in these ranges, the RMSE (root mean square error) of ϕ_4 with $N = 10$ may be as small as some of the RMSE of ϕ_1 with N as large as hundreds (impractical). The bias of ϕ_4 increases when true ϕ goes to the bounds of the ranges of "low ϕ" or "high ϕ" and becomes quite large when ϕ is beyond the ranges. However, due to small SD, ϕ_4 is still a reasonable estimate if ϕ is not too far beyond.

For non-zero $\eta(t)$, we create $\eta(t)$ following two types of models for non-stationary series combined with two ARMA model specifications and with three different levels of signal-to-noise ratio. We found that, in most situations the signal-to-noise ratio has no significant impact on the SD but in some situations, the SD may increase with signal-to-noise ratio. "Mean Diff." decreases when signal-to-noise ratio increases and high signal-to-noise ratio makes ϕ_4 concentrating somewhere in the core range of "low ϕ" and towards the lower bound of "high ϕ" in despite of the true value of ϕ in the corresponding range.

Appendix A. Some histograms of A

Table A1 Some histograms of A

	ϕ								
Interval	0.00	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
(3.06,3.14)	0	0	0	0	0	0	0	1	0
(2.98,3.06)	0	0	0	0	0	0	1	15	28
(2.91,2.98)	0	1	1	0	3	6	27	89	178
(2.83,2.91)	0	0	2	6	10	25	90	304	470
(2.75,2.83)	1	4	5	15	24	90	217	560	847
(2.67,2.75)	0	5	18	29	79	168	416	944	1122
(2.59,2.67)	0	17	31	52	108	269	657	1086	1355
(2.51,2.59)	4	29	33	91	219	440	821	1277	1247
(2.43,2.51)	3	44	90	161	330	648	991	1198	1150
(2.36,2.43)	27	76	134	291	493	746	1046	1067	923
(2.28,2.36)	27	140	243	402	605	877	1050	906	689
(2.20,2.28)	57	248	391	542	780	996	966	712	566
(2.12,2.20)	110	371	515	719	907	1038	875	564	406
(2.04,2.12)	223	535	682	837	1007	956	782	424	325
(1.96,2.04)	406	714	918	1078	1069	880	620	250	195
(1.88,1.96)	583	1047	1098	1084	982	764	456	208	147
(1.81,1.88)	961	1239	1186	1059	859	632	338	125	101
(1.73,1.81)	1454	1247	1123	980	719	474	227	86	84
(1.65,1.73)	1649	1185	1028	799	600	361	139	56	45
(1.57,1.65)	1553	1010	848	610	426	214	104	44	37
(1.49,1.57)	1234	786	587	466	300	151	62	40	25
(1.41,1.49)	752	522	424	305	181	95	37	11	18
(1.34,1.41)	406	316	251	190	107	60	32	16	13
(1.26,1.34)	256	197	167	120	74	43	17	8	10
(1.18,1.26)	119	114	93	61	47	28	15	3	5
(1.10,1.18)	72	68	54	44	27	9	5	2	4
(1.02,1.10)	40	37	36	15	19	19	7	2	5
(0.94,1.02)	31	26	19	21	12	4	1	0	1
(0.86,0.94)	10	6	13	19	8	6	0	0	0
(0.79,0.86)	6	12	6	3	4	1	1	2	1
(0.71,0.79)	3	2	3	1	1	0	0	0	0
(0.63,0.71)	2	1	0	0	0	0	0	0	1
(0.55,0.63)	0	1	1	0	0	0	0	0	2
(0.47,0.55)	1	0	0	0	0	0	0	0	0
(0.39,0.47)	0	0	0	0	0	0	0	0	0
Mean	1.68	1.77	1.83	1.91	2.00	2.13	2.28	2.43	2.48
SD	0.22	0.27	0.28	0.30	0.31	0.31	0.30	0.27	0.28

Note: The sum of numbers in each column is 10,000. The length of interval= $\pi/40$.

Appendix B. Steps of getting ϕ_2 for "high ϕ"

On one hand, with some probability, we may obtain a value A which is much lower than the mean value (see Table 3.4 or Table A1) for a very high ϕ. Then, simply using A like (3.12) (see Step 3 in the following) to get ϕ_2 will have large impact on $G(\phi_2)$. In fact, for a very high ϕ, small SC is a very strong feature which may rectify the wrong information given by A. Step 1 in the following is for this purpose. On the other hand, for testing "high ϕ" or "low ϕ", we suggest to set "$H_0 : \phi \geq 0.7$", i.e. the null hypothesis is "high ϕ", and take $A_C = 1.96$. Then the chance of mis-identifying a very high ($\phi > 0.8$) as "low ϕ" is small (see Table 3.5). However, the power for $0.3 < \phi < 0.7$ is low. For example, when $\phi = 0.5$, the power is 0.58, and hence the probability of mis-identifying $\phi = 0.5$ as "high ϕ" is 0.42. If this mistake occurs, we should try to obtain a relatively low ϕ_2 (says, less than 0.8), then the impact on $G(\phi_2)$ is small (see Table 3.3). In other words, we should avoid to assign a very high ϕ_2 to a false "high ϕ" or to a "not very high ϕ". Step 2 in the following may help. Roughly speaking, the first two steps revise ϕ_1 based on SC. The threshold numbers in these steps are summarized from our simulation results. Step 3 gives an estimate of ϕ totally based on A, and ϕ_2 is defined in Step 4 by averaging both.

Step 1. This step is getting $\phi_2(1)$ to raise ϕ_1 which could be too low for a higher ϕ. Let $\phi_2(1) = \max\{j, \phi_1\}$, where j takes the following values when SC is in the range inside the brackets (Note that $\phi_2(1) = \phi_1$ if $SC > 20$):

$$0.9\ (SC \leq 3), \quad 0.85\ (3 < SC \leq 5), \quad 0.8\ (5 < SC \leq 10),$$
$$0.7\ (10 < SC \leq 15), \quad 0.6\ (15 < SC \leq 20), \quad 0\ (SC > 20).$$

Step 2. This step provides a further adjustment $\phi_2(2)$ from $\phi_2(1)$. In fact, it may pull down ϕ_1 which is too high for a lower ϕ. Note that if ϕ_1 has already been raised ($\phi_2(1) \geq \phi_1$), it will not be pulled down. $\phi_2(2)$ takes the value as follows when SC falls in the corresponding brackets:

For $\phi_2(1) \geq 0.9$:

$\phi_2(1)$ $(SC \leq 10)$, 0.9 $(10 < SC \leq 15)$, 0.85 $(15 < SC \leq 20)$,
0.8 $(20 < SC \leq 25)$, 0.7 $(25 < SC \leq 35)$, 0.6 $(SC > 35)$.

For $0.85 \leq \phi_2(1) < 0.9$:

$\phi_2(1)$ $(SC \leq 20)$, 0.85 $(20 < SC \leq 25)$, 0.8 $(25 < SC \leq 35)$,
0.7 $(35 < SC \leq 45)$, 0.6 $(SC > 45)$.

For $0.8 \leq \phi_2(1) < 0.85$:

$\phi_2(1)$ $(SC \leq 35)$, 0.8 $(35 < SC \leq 45)$, 0.7 $(45 < SC \leq 55)$, 0.6 $(SC > 55)$.

For $0.7 \leq \phi_2(1) < 0.8$:

$\phi_2(1)$ $(SC \leq 55)$, 0.7 $(55 < SC \leq 75)$, 0.6 $(SC > 75)$.

For $0.6 \leq \phi_2(1) < 0.7$:

$\phi_2(1)$ $(SC \leq 75)$, 0.6 $(SC > 75)$.

For $\phi_2(1) < 0.6$: $\phi_2(2) = \phi_2(1)$.

Step 3. Similar to (3.12), let

$$\phi_2(3) = \begin{cases} 0.2, & \text{if } A < M_{0.3}(A) = 1.77, \\ j, & \text{if } M_j(A) \leq A < M_{j+0.1}(A), \quad j = 0.3, \cdots, 0.8; \\ 0.9, & \text{if } M_j(A) \leq A < M_{0.95}(A); \\ 0.95 & \text{if } A \geq M_{0.95}(A) = 2.48. \end{cases}$$

Step 4.

$$\phi_2 = (\phi_2(2) + \phi_2(3))/2.$$

Appendix C. The estimation of σ^2

Now, ϕ_4 is considered to be the best estimate of ϕ. Denote the root mean square error (RMSE) of ϕ_4 by δ (it may depend on ϕ). Then the estimate σ_4^2 of σ^2 based on ϕ_4 is obtained by minimizing [c.f. (2.17), we take $K = 25$]

$$S(\phi_4, \sigma^2) = \sum_{k=14}^{K} \{\hat{v}_{e^*\varepsilon}(k) - \rho(k, \phi_4)\sigma^2/(1 - \phi_4^2)\}^2 \tag{C.1}$$

in the interval $\Sigma_4 = [\sigma_{4l}^2, \sigma_{4u}^2]$ where [c.f. (2.18) and (2.19)]

$$\sigma_{4l}^2 = \hat{v}_e(0)(1 - p\%)(1 - \phi_{4u}^2), \quad \sigma_{4u}^2 = \hat{v}_e(0)(1 + p\%)(1 - \phi_{4l}^2), \tag{C.2}$$

$\hat{v}_e(0)$ is the estimate of $v_e(0)$ published by statistical agencies with $v_e(0) \in [\hat{v}_e(0)(1 - p\%), \hat{v}_e(0)(1 + p\%)]$. In (C.2), for "low ϕ",

$$\phi_{4l} = \max\{0.3, \phi_4 - \delta\}, \quad \phi_{4u} = \min\{0.8, \phi_4 + \delta\}; \tag{C.3}$$

and for "high ϕ",

$$\phi_{4l} = \max\{0.5, \phi_4 - \delta\}, \quad \phi_{4u} = \min\{0.95, \phi_4 + \delta\}. \tag{C.4}$$

This means that the two overlapped intervals [0.3, 0.8] and [0.5, 0.95] are assumed to be the ranges of "low ϕ" and "high ϕ" respectively.

The following discussion gives some idea of what Σ_4 could be. Since δ is the RMSE, we may assume the first situation is that ϕ falls in (ϕ_{4l}, ϕ_{4u}) with probability about 2/3. If so, we can show that $\sigma^2 \in \Sigma_4$; i.e. the minimization range (for getting σ_4^2) covers the true value of σ^2. The second situation is that, with probability about 1/3, ϕ falls outside (ϕ_{4l}, ϕ_{4u}), but $\delta \leq |\phi_4 - \phi| \leq 2\delta$. In this situation, we can show that Σ_4 is inside either $[\sigma_l^2(3\delta), \sigma^2]$ or $[\sigma^2, \sigma_u^2(3\delta)]$, where

$$\sigma_l^2(\delta) = v_e(0)/(1 - \phi_u(\delta)^2), \quad \sigma_u^2(\delta) = v_e(0)/(1 - \phi_l(\delta)^2); \tag{C.5}$$

and for "low ϕ",

$$\phi_l(\delta) = \max\{0.3, \phi - \delta\}, \quad \phi_u(\delta) = \min\{0.8, \phi + \delta\}; \tag{C.6}$$

for "high ϕ",

$$\phi_l(\delta) = \max\{0.5, \phi - \delta\}, \quad \phi_u(\delta) = \min\{0.95, \phi + \delta\}. \tag{C.7}$$

In the second situation, $|\phi - \phi_4| = 2\delta$ is the worst; if that happens, Σ_4 keeps away from σ^2 by either $\sigma_l^2(\delta)$ or $\sigma_u^2(\delta)$. Table C1 shows some values of these bounds for "high ϕ" when $\sigma^2 = 1$.

Table C1 Some values of $\sigma_l^2(.)$ and $\sigma_u^2(.)$ for true $\sigma^2 = 1$

ϕ	0.5	0.6	0.7	0.8	0.9	0.95
$\sigma_l^2(3\delta)$	0.47	0.29	0.19	0.27	0.51	1
$\sigma_l^2(\delta)$	0.85	0.80	0.70	0.53	0.51	1
$\sigma_u^2(\delta)$	1	1.17	1.25	1.42	1.89	2.84
$\sigma_u^2(3\delta)$	1	1.17	1.47	2.09	3.37	5.92

Table C2 Statistics of σ_4 for "low ϕ", $\sigma = 1, \delta = 0.08$

	ϕ	0.3	0.4	0.5	0.6	0.7	0.8
Mean	p=2	0.852	0.885	0.934	1.005	1.114	1.296
	p=20	0.836	0.874	0.927	1.014	1.115	1.300
	p=50	0.802	0.845	0.905	0.987	1.104	1.289
RMSE	p=2	0.156	0.127	0.090	0.070	0.140	0.313
	p=20	0.182	0.154	0.121	0.107	0.168	0.334
	p=50	0.239	0.213	0.185	0.174	0.221	0.369
SD	p=2	0.051	0.055	0.061	0.070	0.081	0.101
	p=20	0.081	0.088	0.096	0.107	0.122	0.148
	p=50	0.135	0.146	0.158	0.174	0.195	0.229

Table C3 Statistics of σ_4 for "high ϕ", $\sigma = 1, \delta = 0.08$

	ϕ	0.5	0.6	0.7	0.8	0.9	0.95
Mean	p=2	0.707	0.750	0.805	0.883	0.984	1.103
	p=20	0.694	0.740	0.799	0.880	0.976	1.080
	p=50	0.667	0.718	0.781	0.865	0.964	1.052
RMSE	p=2	0.323	0.294	0.268	0.257	0.273	0.226
	p=20	0.338	0.308	0.280	0.269	0.285	0.228
	p=50	0.372	0.340	0.309	0.290	0.290	0.241
SD	p=2	0.135	0.153	0.184	0.229	0.273	0.201
	p=20	0.144	0.166	0.196	0.240	0.284	0.213
	p=50	0.166	0.189	0.218	0.256	0.288	0.236

Table C2 and C3 show some statistics of σ_4 (not σ_4^2) for the same 10,000 replications. The corresponding ϕ_4 has statistics listed in Table 4.1 and 4.2. The true $\sigma = 1$; we take $\delta = 0.08$ where most RMSE of ϕ_4 are around this value. If we use $\delta = 0.2$, the results are much worse; if we use $\delta = 0.05$, the results are slightly better for $\phi \leq 0.9$

but worse for $\phi = 0.95$. Thus a value between 0.05 and 0.08 is recommended.

The RMSE of σ_4 is larger for larger p. This effect is more obvious for "low ϕ" than for "high ϕ" since SD increases drastically for "low ϕ". With the same p, the higher the ϕ, the larger the RMSE. That is because fixing $\sigma^2 = 1$, $v_e(0) = 1/(1-\phi^2)$ is larger for higher ϕ, so $\hat{v}_e(0)$ is usually larger with higher ϕ and the range for minimization [see (C.2)] becomes larger which probably causes a larger error.

In the core ranges of either "low ϕ" or "high ϕ", σ_4 is almost unbiased as ϕ_4 is almost unbiased. Over-estimation of ϕ leads to under-estimation of σ, and vice verse.

Appendix D. The choice of parameters for trimming $\hat{\rho}(k)$

Now we demonstrate that the choices of ϕ_{min}, ϕ_{max} and α in (3.6) and (4.1) have great impact on the estimate. It affects the unbiasedness of ϕ_4 in the core ranges of "low ϕ" or "high ϕ" where the bias is expected to be the least. It also affects the SD. Table D1 and D2 list the values of the mean and the RMSE of ϕ_4 for some true values of ϕ (in the top rows of the tables) with different choices of ϕ_{min}, ϕ_{max} and α based on 10,000 replications. The range of the SD of ϕ_4 for ϕ taking these listed values are in the last columns of the tables.

For "low ϕ" (Table D1), as $(\phi_{min}, \phi_{max}) = (0.4, 0.7)$ gives the smallest SD for $0.3 \leq \phi \leq 0.8$ and the smallest RMSE for $0.4 \leq \phi < 0.7$, we choose $(\phi_{min}, \phi_{max}) = (0.4, 0.7)$ and obtain Table 4.1 although $(\phi_{min}, \phi_{max}) = (0.3, 0.8)$ leads to less biased ϕ_4.

Note that ϕ_4 in Table D1 is obtained from letting $\phi_2 =$ true ϕ. The statistics in the rows of $(\phi_{min}, \phi_{max}) = (0.4, 0.7)$ is slightly different from those of ϕ_4 in Table 4.1 where ϕ_2 is obtained from A. This explains our procedure of obtaining a preliminary ϕ_2 from A is good enough for getting final estimate of ϕ_4.

Table D1 Statistics of ϕ_4 for "low ϕ" with different ϕ_{min} and ϕ_{max}

	ϕ_{min}	ϕ_{max}	0.3	0.4	0.5	0.6	0.7	0.8	SD Range
Mean	0.3	0.7	0.495	0.509	0.526	0.549	0.579	0.622	0.088-0.100
	0.3	0.8	0.529	0.547	0.567	0.592	0.628	0.677	0.099-0.107
	0.4	0.7	0.542	0.553	0.566	0.583	0.607	0.644	0.063-0.084
	0.4	0.8	0.575	0.589	0.604	0.625	0.654	0.696	0.073-0.092
RMSE	0.3	0.7	0.214	0.141	0.092	0.102	0.150	0.204	
	0.3	0.8	0.250	0.177	0.120	0.100	0.123	0.163	
	0.4	0.7	0.251	0.165	0.091	0.068	0.116	0.178	
	0.4	0.8	0.285	0.202	0.128	0.081	0.093	0.139	

Table D2 Statistics of ϕ_4 for "high ϕ" with different ϕ_{min} and $\alpha, \phi_{max} = 0.9$

	ϕ_{min}	α	0.5	0.6	0.7	0.8	0.9	0.95	SD Range
Mean	0.5	1.0	0.637	0.653	0.677	0.714	0.789	0.873	0.055-0.112
	0.5	1.5	0.725	0.748	0.777	0.819	0.882	0.932	0.076-0.085
	0.5	2.0	0.786	0.811	0.846	0.881	0.929	0.961	0.052-0.092
	0.4	1.5	0.689	0.716	0.753	0.800	0.870	0.927	0.085-0.099
	0.3	1.5	0.655	0.688	0.731	0.785	0.861	0.922	0.098-0.199
RMSE	0.5	1.0	0.148	0.081	0.072	0.121	0.155	0.136	
	0.5	1.5	0.238	0.166	0.110	0.083	0.087	0.078	
	0.5	2.0	0.300	0.229	0.165	0.109	0.071	0.053	
	0.4	1.5	0.213	0.152	0.111	0.096	0.101	0.088	
	0.3	1.5	0.199	0.151	0.121	0.112	0.116	0.098	

For "high ϕ", for simplicity, we always use $\phi_{max} = 0.9$ as $\rho(k, 0.9)$ is almost in the lowest position for $k \leq 20$. Although for $k = 14,\ 15$, $\rho(k, 0.8)$ is lower, but we may ignore that as we finally use $\alpha > 1$. As the Mean is concerned, from Table D2, the best choice is $\alpha = 1.5$ with $\phi_{min} = 0.5$ or 0.4. According to the RMSE, 0.5 is better than 0.4 in $[0.7, 0.95]$ (the supposed range of "high ϕ"). Taking $\alpha = 1$ is good for $\phi \leq 0.7$ but too bad for $\phi \geq 0.8$ (the core range of "high ϕ"). $\alpha = 2$ is good for $\phi \geq 0.9$, but not good for $\phi \leq 0.8$.

Appendix E. Some further comments on the results in Table 4.1 and 4.2

1. The RMSE of ϕ_3 or ϕ_4 is somewhat larger at the bounds (0.7 for "high ϕ", 0.4 and 0.7 for "low ϕ"). It seems unfair for ϕ in the neighbourhood of 0.7. But notice that, with some probability, $\phi = 0.6$ may be mis-identified as "high ϕ" and get even larger RMSE due to over-estimated. Similar problem occurs for $\phi = 0.8$ if it is mis-identified as "low ϕ". That means the RMSE showed in Table 4.1 and 4.2 are the "conditional" RMSE (0.6 or 0.8 is correctly identified to "low ϕ" or "high ϕ"). The unconditional one will be larger. But whatever 0.7 is regarded as "low ϕ" or "high ϕ", the RMSE is always around $0.11 \sim 0.12$ which is not bad.

2. The RMSE of ϕ_4 becomes larger when ϕ reaches and goes beyond the bounds, but its SD remains unchanged (For "low ϕ", it is about 0.06; for "high ϕ", it is about 0.08). That means the increase of the RMSE is caused by the bias which can also be seen from the Mean in Table 4.1 and 4.2.

3. Mis-identifying a "low ϕ" as "high ϕ" leads to over-estimation and vice versa. Therefore a good preliminary judgment is crucial; but a second look after estimation is sometimes useful. For extended range of "high ϕ" ($0.5 \leq \phi \leq 0.8$), the mean of $\phi_4 - \phi_2$ is $0.09 \sim 0.096$ and this difference is much larger than that for $\phi > 0.8$. If we get low values of ϕ_2 and ϕ_4 for a "high ϕ" (ϕ_4 is less or not much higher than 0.7), also $\phi_4 - \phi_2$ is large, says > 0.1, then we should think the true ϕ is a "low ϕ". For extended "low ϕ", the mean of $\phi_4 - \phi_2$ is 0.177, 0.147 and 0.107 for $\phi = 0.3$, 0.4 and 0.5, while it is 0.002 and -0.056 for $\phi = 0.7$ and 0.8. Therefore, a large increase of ϕ_4 from ϕ_2 indicates a very low ϕ. If says, $\phi_2 \leq 0.4$ and $\phi_4 - \phi_2 > 0.15$, we should not use ϕ_4 as the final estimate of ϕ, but use ϕ_2 or $(\phi_2 + \phi_4)/2$.

Appendix F. Some further comments on the results in Table 4.3 and 4.4

1. For saving space in these tables, we did not list the means of ϕ_4 for each ϕ. However the RMSE may roughly explain that: Since the SD does not change too much with ϕ in the same situation, the minimum of the RMSE indicates the neighbourhood of ϕ where ϕ_4 is almost unbiased. For "low ϕ", the unbiasness reaches at a value of ϕ roughly between 0.5 and 0.6 in all situations. Any ϕ higher than this value is under-estimated and any ϕ lower than this value is over-estimated. Larger "Mean Diff." implies this under- or over-estimation is less serious. For "high ϕ", the bias of ϕ_4 may reach minimum at the true ϕ located mostly between 0.8 and 0.95 depending on situations. For low signal-to-noise ratio, it is usually close to 0.95 and moves towards 0.8 while the signal-to-noise ratio increases. In general, the "Mean Diff." is relatively much larger for "high ϕ" than that for "low ϕ". The largest "Mean Diff." in Table 4.4 is 0.166 (when $\eta(t) = 0$) corresponding to the true difference $0.95 - 0.7 = 0.25$ while the largest "Mean Diff." in Table 4.3 for "low ϕ" is 0.057 corresponding to the true difference 0.7-0.4=0.3.

2. We may increase "Mean Diff." by increasing ϕ_{max} - ϕ_{min}. For example, using (0.3, 0.8) instead of (0.4, 0.7) as (ϕ_{min}, ϕ_{max}). From Table D1, we may see that "Mean Diff." increases from 0.054 to 0.081, however the SD also increases by about 0.03 if (0.3, 0.8) are used to replace (0.4, 0.7).

3. When σ_η increases, "Mean Diff." decreases with different rate depending on situations. The decrease is more severe for TS + MA and DS + AR. The "Mean Diff." is small or very small with $\sigma_\eta = 3$, especially for the situations TS + MA and DS + AR. It is conceivable that $\sigma_\eta = 3$ makes a situation very difficult to deal with as the disturbance from the variable (a nuisance) is much stronger than the survey error in which we are interested. In such a situation, taking a value as an estimate for a "low ϕ" or a "high ϕ" around the middle of a supposed interval is a compromise between "making good decision" and "lack

of information". The bias of ϕ_4 for those ϕ close to one or both bounds of the range may be relatively large, but the SD of ϕ_4 is as small as for other ϕ.

4. When σ_η increases from 0, there is no substantial change of the SD for most situations except DS+AR. At that situation, even $\sigma_\eta = 1/3$ makes the SD much larger than those of $\sigma_\eta = 0$.

ACKNOWLEDGEMENT

The authors thank Marietta Morry for her strong support of this work and thank Benoit Quenneville, Wai-Sum Chan, Siu-Hung Cheung, Pierre Chollete and John Higginson for many valuable comments, helping with translating the abstract to French or with some editing. This research was partially supported by the Physical Science Panel Direct Grant for Research of the Chinese University of Hong Kong.

REFERENCES

Bell, W. R. and Wilcox, D. W. (1993). The effect of sampling error on time series behavior of consumption data. *Journal of Economics*, **55**, 235-265.

Chen, Z-G., Cholette, P. A. and Dagum, E. B. (1997). A nonparametric method for benchmarking survey data via signal extraction. *Journal of American Statistical Association*, **92**, 1563-1571.

Chollete, P. A. and Dagum, E. B. (1994). Benchmarking time series with autocorrelated survey errors. *International Statistical Review*, **62**, 365-377.

Denton, F. T. (1971). Adjustment of monthly or quarterly series to annuals: An approach based on quadratic minimization. *Journal of American Statistical Association*, **66**, 99-102.

Durbin, J. and Quenneville, B. (1997). Benchmarking by state space models. *International Statistical Review*, **65**, 23-48.

Hillmer, S. C. and Trabelsi, A. (1987). Benchmarking of economic time series. *Journal of American Statistical Association*, **82**, 1064-1071.

Lee, H. (1990). Estimation of panel correlation for the Canadian labour force survey. *Survey Methodology*, **16**, 283-292.

Nelson, C. R. and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series. *Journal of Monetary Economics*, **10**, 139-162.

Pfeffermann, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys. *Journal of Business and Economic Statistics*, **9**, 163-175.

Pfeffermann, D. and Scott, S. (1997). Variance measures for X-11 seasonally adjusted estimators; some new developments with application to labor force series. *Proceedings of the ASA Section on Business and Economic Statistics*, 211-216.

Scott, A. J., Smith, T. M. F. and Jones, R. G. (1977). The application of time series methods to the analysis of repeated surveys. *International Statistical Review*, **45**, 13-28.

Trabelsi, A. and Hillmer, S. T. (1990). Bench-marking time series with reliable bench-marks. *Applied Statistics (JRSS, Series C)*, **39**, 367-379.

STATISTICS CANADA LIBRARY
BIBLIOTHÈQUE STATISTIQUE CANADA
1010329777
c. 2