Statistics Canada Statistique Canada

Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes-auprès des entreprises

Canada

COMPARISON OF BENCHMARKING METHODS WITH AND WITHOUT SURVEY ERROR MODELLING

C.2

By

Zhao-Guo Chen and Ka Ho Wu

BSMD - 2003 - 006E

COMPARISON OF BENCHMARKING METHODS WITH AND WITHOUT SURVEY ERROR MODELLING

by

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

November 2003
Time Series Research and Analysis Centre
Business Survey Methods Division
Statistics Canada

[1]Time Series Centre, 17H R.H.Coats Building, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6 (chenzha@statcan.ca)

[2]Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong (wu@sta.cuhk.edu.hk)

ABSTRACT

For a target socio-economic variable, two sources of data with different precisions and collecting frequencies may be available. Typically, the less frequent data (e.g. annual report or census) are more reliable and are considered as benchmarks. The process of using them to adjust the more frequent and less reliable data (e.g. repeated monthly survey) is called benchmarking. For the implementation of some advanced benchmarking procedures, the survey error model is needed and usually is not given. In this paper, we will show the relationship among three types of benchmarking methods in the literature, namely the Denton (original and modified), the regression, and the signal extraction methods. Assuming the survey error series follows an AR(1) model, by simulation, we investigate the impact of mis-specification of the model on the benchmarking predictions based on the criterion of minimizing root-mean-squared error of prediction. It is concluded that the survey error modelling procedure proposed by Chen and Wu (2000, 2001) may lead to results as good as those obtained from using the true survey error models.

Résumé

Deux sources de données de précisions différentes et avec des fréquence de collecte différentes peuvent être disponibles pour une variable socio-économique. D'habitude, la source de données moins fréquente, par exemple un rapport annuel ou bien un recensement, est plus fiable et est considérée comme un étalon. On appelle étalonnage le processus qui consiste à utiliser la source de données moins fréquente pour corriger les données plus fréquentes, par exemple une enquête mensuelle répétée. Pour la mise en œuvre de certaines méthodes d'étalonnage, le modèle pour l'erreur d'échantillonnage est nécessaire mais est rarement disponible. Dans cet article nous examinons la relation entre trois méthodes d'étalonnage, à savoir la méthode de Denton (originale et modifiée), la méthode basée sur un modèle de régression, et la méthode basée sur l'extraction du signal. En supposant un que l'erreur d'échantillonnage suit un processus autorégressif d'ordre 1, nous examinons, à l'aide d'une simulation, l'impact d'une mauvaise spécification du modèle sur les données étalonnées en utilisant un critère de minimisation de l'erreur quadratique moyenne des prévisions. Nous concluons que la méthode de modélisation de l'erreur d'échantillonnage proposée par Chen et Wu (2000, 2001) peut donner des résultats aussi bons que ceux obtenus en utilisant le vrai modèle pour l'erreur d'échantillonnage.

1. Introduction

For a target socio-economic variable, two sources of data with different precisions and collecting frequencies may be available. Typically, the less frequent data (e.g. annual report or census) are more reliable and considered as benchmarks. The process of using them to adjust the more frequent and less reliable data (e.g. repeated monthly survey) is called benchmarking. A good example of benchmarking is the adjustment of the monthly retail trade series obtained from surveys by the annual total figures obtained from a more reliable source, e.g. see Hillmer and Trabelsi (1987), and Dagum, Chollete, and Chen (1998). Simulations in Chen, Cholette, and Dagum (1997) showed that by using some advanced benchmarking methods, the root-mean-squared error of prediction of the target variable may be reduced by more than 40% from the currently used methods.

Suppose we have monthly observations $y(t)$

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

where $\eta(t)$ is the target socio-economic variable and $e(t)$ is the monthly survey error. Also, suppose that we have the annual sum, the benchmarks $z(T)$, obtained from a more reliable source, i.e.

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

where the notation $t \in T$ means month t is in year T. The benchmarking problem is then to predict $\eta(t)$ using both the monthly survey data $y(t)$ and the annual benchmarks $z(T)$. The predictions of $\eta(t)$ are called the benchmarking predictions or the benchmarked values. In this paper, we consider the situation that all $z(T)$ do not contain observation errors as described in (1.2). In this case, $z(T)$ are called binding benchmarks; otherwise, $z(T)$ are called non-binding benchmarks. We assume that the survey error series follows the model

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

where $0 \leq \phi \leq 1$ and $\{\xi(t)\}$ is a white noise series with mean zero and variance σ^2. It is a stationary AR(1) model when $0 \leq \phi < 1$ and a random walk when $\phi = 1$. It was pointed out by many authors [e.g., Scott, Smith and Jones (1977), and Chen and Wu (2001)] that this assumption is reasonable when the survey design does not involve a complicated panel rotation.

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

$$\begin{cases} \mathbf{y} &= \boldsymbol{\eta} + \mathbf{e} \\ \mathbf{z} &= \mathbf{L}\boldsymbol{\eta} \end{cases} \tag{1.4}$$

where $\mathbf{y} = (y(1), \cdots, y(n))'$, $\boldsymbol{\eta} = (\eta(1), \cdots, \eta(n))'$, $\mathbf{z} = (z(1), \cdots, z(N))'$ etc., $n \geq 12N$. $\mathbf{L}$ is a matrix of 0's and 1's which relates the monthly values to the benchmarks. For example, if $n = 12N$, then

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

Denton (1971) introduced a well-known method where the benchmarking prediction $\hat{\boldsymbol{\eta}} = (\hat{\eta}(1), \cdots, \hat{\eta}(n))'$ of $\boldsymbol{\eta}$ minimizes the penalty function

$$p(\boldsymbol{\eta}, \mathbf{y}) = (\boldsymbol{\eta} - \mathbf{y})'\mathbf{A}(\boldsymbol{\eta} - \mathbf{y}) \tag{1.5}$$

subject to the constraint $\mathbf{z} = \mathbf{L}\boldsymbol{\eta}$ for a reasonable choice of a symmetric $n \times n$ positive definite matrix $\mathbf{A}$. Thus we have

$$\hat{\boldsymbol{\eta}} = \mathbf{y} + \mathbf{A}^{-1}\mathbf{L}'(\mathbf{L}\mathbf{A}^{-1}\mathbf{L}')^{-1}(\mathbf{z} - \mathbf{L}\mathbf{y}). \tag{1.6}$$

Denton (1971) suggested keeping the month-to-month changes as small as possible [equivalently the movement of $\hat{\eta}(t)$ as close as possible to that of $y(t)$]. Thus the penalty function (1.5) becomes

$$\begin{aligned} p(\boldsymbol{\eta}, \mathbf{y}) &= \textstyle\sum_{t=1}^{n}\{[\eta(t) - \eta(t-1)] - [y(t) - y(t-1)]\}^2 \\ &= \textstyle\sum_{t=1}^{n}\{[\eta(t) - y(t)] - [\eta(t-1) - y(t-1)]\}^2, \end{aligned} \tag{1.7}$$

with initial value $\eta(0) = y(0)$. As a result, $\mathbf{A} = \mathbf{P}'\mathbf{P}$ where

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & .. & 0 & 0 \\ -1 & 1 & 0 & .. & 0 & 0 \\ 0 & -1 & 1 & .. & 0 & 0. \\ . & . & . & .. & . & . \\ 0 & 0 & . & .. & -1 & 1 \end{pmatrix}, \tag{1.8}$$

and

$$\mathbf{A}^{-1} = \begin{pmatrix} 1 & 1 & .. & 1 \\ 1 & 2 & .. & 2 \\ . & . & .. & . \\ 1 & 2 & .. & n \end{pmatrix}. \tag{1.9}$$

Cholette (1984) noted that spurious fluctuations of $\hat{\eta}(t)$ at the beginning of the series often occur due to the imposed initial condition $\eta(0) = y(0)$. Consequently, he suggested using the following modified penalty function

$$p(\eta, \mathbf{y}) = \sum_{t=2}^{n} \{[\eta(t) - y(t)] - [\eta(t-1) - y(t-1)]\}^2. \tag{1.10}$$

This modification is well received by many practitioners as it keeps the early part of the series $\hat{\eta}(t)$ having a similar movement to $y(t)$ and provides smooth backcasts. The modified Denton method is widely used by many statistical agencies. However, the matrix $\mathbf{A}$ corresponding to (1.10) becomes a $(n-1) \times n$ matrix obtained by deleting the first row of the matrix in (1.8). Thus $\mathbf{A} = \mathbf{P}'\mathbf{P}$ is degenerate and the formula (1.6) no longer works and another more complicated algorithm must be used (Cholette, 1984).

From the statistical point of view, one would prefer that the benchmarking predictions minimize the variances of the prediction errors, $\mathrm{Var}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$. The Denton method did not address the problem in such a way. This drawback has been well recognized by many authors including Hillmer and Trabelsi (1987), Cholette and Dagum (1994), and Chen, Cholette and Dagum (1997). Consequently, several advanced benchmarking methods have been derived. However, all these methods, including the state-space approach (Durbin and Quenneville, 1997), require the autocorrelation of the survey error, or equivalently, its time series model. Unfortunately, some very restrictive and unrealistic assumptions for the model of $\eta(t)$ are required. See for example, Scott, Smith and Jones (1977).

Assuming the covariance matrix $\mathbf{V}_e$ of the survey error $\mathbf{e}$ is known, Cholette and Dagum (1994) introduced a benchmarking method based on regression, regarding (1.4) as a regression model with "parameters" $\boldsymbol{\eta}$ and errors $\mathbf{e}$, where $e(t)$ is a stationary series. If the mean of $e(t)$ is zero, then the benchmarked value $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$ which minimizes $\mathrm{Var}(\hat{\boldsymbol{\eta}}$ -

η) is

$$\hat{\eta} = \mathbf{y} + \mathbf{V_e L}'(\mathbf{L V_e L}')^{-1}(\mathbf{z} - \mathbf{Ly}). \tag{1.11}$$

This is the generalized least squares (GLS) solution for the regression model and $\hat{\eta}$ is the best linear unbiased estimate (BLUE) of η.

Comparing (1.11) with (1.6), we see that the benchmarking formula is the same as that of the original Denton except that $\mathbf{A}^{-1}$ in (1.6) is now replaced by $\mathbf{V_e}$.

Assuming $e(t)$ follows model (1.3), $0 \leq \phi < 1$, the variance and the autocovariances of $e(t)$ are

$$v_e(k) = \phi^k \sigma^2/(1-\phi^2) = v_e(0)\phi^k, \quad k = 0, 1, 2, \cdots. \tag{1.12}$$

Then $\mathbf{V_e}$ in (1.11) can be replaced by

$$\mathbf{\Phi} = \begin{pmatrix} 1 & \phi & \ldots & \phi^{n-1} \\ \phi & 1 & \ldots & \phi^{n-2} \\ \ldots & \ldots & \ldots & \ldots \\ \phi^{n-1} & \phi^{n-2} & \ldots & 1 \end{pmatrix}.$$

Assume $e(t)$ is a random walk [Without loss of generality, we set $\sigma^2 = 1$]; then $\mathbf{A}^{-1}$ in (1.9) is the conditional covariance matrix of $\mathbf{e} = (e(1), \cdots, e(n))'$ given $e(0)$. Some authors tried to build consistency of Denton's solution with the BLUE. Fernandez (1981) assumed $e(0) = 0$ and left-multiplied the first equation of (1.4) by $\mathbf{P}$ [$\mathbf{P}$ is given by (1.8)]; then the BLUE of η becomes (1.6). Note that $e(0)$ is unknown in practice. Due to (1.3), rewrite (1.1) as $y(t) = e(0) + \eta(t) + e'(t)$, where $e'(t) = \sum_{j=1}^{t} \xi(j)$, we see that the unknown "parameter" $e(0)$ can be regarded as the "bias" of the survey error in the regression model (see Cholette and Dagum, 1994) and can be estimated together with $\eta(t)$ by the GLS method. This GLS solution is the BLUE of $e(0)$ and η; it should not be the same as Fernandez's (or Denton's) solution. This contradiction is caused by the unreal assumption $e(0) = y(0) - \eta(0) = 0$. Therefore (1.6) does not provide the BLUE when the survey error series is a random walk (usually it does not happen in practice).

In fact, the penalty function formulae based on the original Denton, the modified Denton and the regression methods can be viewed as special cases of the following more general penalty function:

$$p_{\phi,\beta}(\boldsymbol{\eta},\mathbf{y}) = \{\beta[\eta(1) - y(1)]\}^2 + \sum_{t=2}^{n}\{[\eta(t) - y(t)] - \phi[\eta(t-1) - y(t-1)]\}^2. \tag{1.13}$$

Analogous to (1.8), we have

$$\mathbf{P}_{\phi,\beta} = \begin{pmatrix} \beta & 0 & 0 & .. & 0 & 0 \\ -\phi & 1 & 0 & .. & 0 & 0 \\ 0 & -\phi & 1 & .. & 0 & 0. \\ . & . & . & .. & . & . \\ 0 & 0 & . & .. & -\phi & 1 \end{pmatrix}, \quad \mathbf{P}_{\phi,\beta}^{-1} = \begin{pmatrix} 1/\beta & 0 & 0 & .. & 0 & 0 \\ \phi/\beta & 1 & 0 & .. & 0 & 0 \\ \phi^2/\beta & \phi & 1 & .. & 0 & 0 \\ . & . & . & .. & . & . \\ \phi^{n-1}/\beta & \phi^{n-1} & \phi^{n-2} & .. & \phi & 1 \end{pmatrix},$$

and

$$\mathbf{A}_{\phi,\beta} = \mathbf{P}'_{\phi,\beta}\mathbf{P}_{\phi,\beta} = \begin{pmatrix} \beta^2+\phi^2 & -\phi & 0 & \cdots & 0 \\ -\phi & 1+\phi^2 & -\phi & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1+\phi^2 & -\phi \\ 0 & \cdots & 0 & -\phi & 1 \end{pmatrix}.$$

The choice of ϕ $(0 \leq \phi \leq 1)$ determines how much the movement of $y(t)$ is kept in $\hat{\eta}(t)$. The two Denton methods intend to keep it mostly by choosing $\phi = 1$. The choice of β $(0 \leq \beta \leq 1)$ determines how much the correction to $y(1)$ will be. The original Denton method minimizes this correction by choosing $\beta = 1$ while the modified Denton method totally ignores that.

When $\phi \uparrow 1$ and $\beta \uparrow 1$, $\mathbf{A}_{\phi,\beta}^{-1} = \mathbf{P}_{\phi,\beta}^{-1}(\mathbf{P}_{\phi,\beta}^{-1})' \to \mathbf{A}^{-1}$ in (1.9) which gives (1.6), the formula of the original Denton method. Let $\beta^2 = 1 - \phi^2$, then the first entry in $\mathbf{A}_{\phi,\beta}$ is 1, and hence $\mathbf{A}_{\phi,\beta}^{-1} = \mathbf{\Phi}$. Replacing $\mathbf{A}^{-1}$ by $\mathbf{\Phi}$ in (1.6), we obtain the formula of the regression method with survey errors following an AR(1) model. In particular, when $\phi \uparrow 1$ (consequently $\beta \downarrow 0$), (1.13) reduces to (1.10), the modified Denton method. Hence, the benchmarking prediction of the modified Denton method can be approximated by the regression method with a coefficient ϕ very close to 1 in the AR(1) model for the survey error. It means that the modified Denton method, which does not enforce $e(0) = 0$, essentially abandons the random walk assumption for survey errors and is a special case of the regression method. However, the impact on the benchmarking predictions from using a ϕ very close to 1 needs a thorough study. This will be examined that by simulation.

The regression method regards $(\eta(1), \cdots, \eta(n))'$ as a set of constants, the parameters in the regression model (1.4). Hillmer and Trabelsi (1987) regarded $\eta(t)$ as a stochastic series and proposed a benchmarking procedure based on signal extraction. [As $\eta(t)$ are regarded as random variables, we call $\hat{\eta}(t)$ "prediction" rather than "estimate".] Chen, Cholette and Dagum (1997) assume $\eta(t)$ follows a "difference stationary" (DS) model:

$$\nabla\nabla_{12}\, \eta(t) = \zeta(t); \tag{1.14}$$

where $\zeta(t)$ is a stationary series with mean zero, and possibly over-differenced, $\nabla = 1 - B$, $\nabla_{12} = 1 - B^{12}$, and B is the backshift operator defined by $B^k\eta(t) = \eta(t-k)$. This is a very general nonstationary model which can fit many real series very well and is widely used. The term "DS" was introduced by Nelson and Plosser (1982). With model (1.14), the benchmarking prediction $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$ via signal extraction is given by

$$\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\eta}}_0 + \hat{\boldsymbol{\eta}}_C, \tag{1.15}$$

where

$$\hat{\boldsymbol{\eta}}_0 = \boldsymbol{\Omega}_0\mathbf{V}_{\mathbf{e}}^{-1}\mathbf{y}, \tag{1.16}$$

$$\hat{\boldsymbol{\eta}}_C = \boldsymbol{\Omega}_0\mathbf{L}'(\mathbf{L}\boldsymbol{\Omega}_0\mathbf{L}')^{-1}(\mathbf{z} - \mathbf{L}\hat{\boldsymbol{\eta}}_0), \tag{1.17}$$

$$\boldsymbol{\Omega}_0 = (\mathbf{V}_{\mathbf{e}}^{-1} + \mathbf{V}_{\boldsymbol{\eta}}^{-1})^{-1}, \tag{1.18}$$

$$\mathbf{V}_{\boldsymbol{\eta}}^{-1} = \mathbf{D}'\mathbf{V}_{\boldsymbol{\zeta}}^{-1}\mathbf{D}, \tag{1.19}$$

and $\mathbf{D}$ is a $(n-13) \times n$ matrix with entries 1, -1 and 0 defined by (1.14). $\mathbf{V}_{\boldsymbol{\zeta}}$ is a $(n-13) \times (n-13)$ Toeplitz matrix with elements $v_\zeta(|i-j|)$, the autocovariances of $\zeta(t)$ at lag $|i-j|$, in its (i,j) entry.

Note that these formulae are of the same format as (1.11): $\mathbf{y}$ in (1.11) is now replaced by $\hat{\boldsymbol{\eta}}_0$, the extracted signal; and $\mathbf{V}_{\mathbf{e}}$ in (1.11), the covariance matrix of $\mathbf{e} = \mathbf{y} - \boldsymbol{\eta}$, is now replaced by $\boldsymbol{\Omega}_0$, the covariance matrix of $\hat{\boldsymbol{\eta}}_0 - \boldsymbol{\eta}$. Regarding $\hat{\boldsymbol{\eta}}_0$ and $\mathbf{y}$ as two different preliminary predictions of $\boldsymbol{\eta}$, $\hat{\boldsymbol{\eta}}_0$ is better than $\mathbf{y}$ (obviously, $\boldsymbol{\Omega}_0 \leq \mathbf{V}_{\mathbf{e}}$). Thus the signal extraction method may provide better results than the regression method. However, for implementing this procedure, we need both $\mathbf{V_e}$ and $\mathbf{V}_{\eta}$.

Both the regression and the signal extraction methods require $\mathbf{V_e}$. Except for $v_e(0)$, the $v_e(k)$ are usually unknown. In Appendix A, we provide a brief discussion of the problem of estimating $\mathbf{V_e}$, or equivalently, of modelling $e(t)$. Moreover, we outline the survey-error-modelling procedure proposed by Chen and Wu (2000, 2001) with a focus on the major steps of the procedure.

In Section 2, we provide a simulation study to compare the performance of the above-mentioned benchmarking methods. The survey-error-modelling procedure suggested by Chen and Wu (2000, 2001) is used and compared with the situations where the parameters of the survey error model are known, either correctly specified (true model) or mis-specified. It concludes that the original Denton method and the modified Denton method are not recommended. It also concludes that for both the regression and the signal extraction methods, the survey error modelling procedure of Chen and Wu may provide predictions as good as those from using the true model.

In practice, $\mathbf{V}_{\zeta}$ is always unknown and is needed in the signal extraction method. The simulation in Section 3 shows that when $\mathbf{V}_{\zeta}$ in (1.19) is replaced by its estimate obtained by the nonparametric method of Chen, Cholette, and Dagum (1997), the impact on the benchmarking prediction is quite large. However the prediction is usually still much better than that from the regression method. The simulation also concludes that the survey-error-modelling procedure of Chen and Wu combined with the nonparametric method for estimating $\mathbf{V}_{\zeta}$ may provide predictions as good as those from using the true survey-error model.

2. Comparison of benchmarking methods

This section provides the simulation study to compare the benchmarking methods in the manner mentioned above. We always assume that the survey error $e(t)$ follows an

AR(1) model as in (1.3) with $0 \leq \phi < 1$. Without loss of generality, we set $\sigma^2 = 1$. We take $\phi = 0.5$ $[v_e(0)^{1/2} = 1.16]$ or $\phi = 0.9$ $[v_e(0)^{1/2} = 2.29]$ as the "true parameter" to generate the data of $e(t)$. These two cases, which represent the survey error series weakly or strongly autocorrelated, are called a "low ϕ" case and a "high ϕ" case in Chen and Wu (2000, 2001).

The target variable $\eta(t)$ is assumed to follow a DS model as in (1.14) with a model specification for $\zeta(t)$ given below. Note that, for the regression method, a specified model for $\zeta(t)$ is unnecessary in the benchmarking formula, but is needed for generating data. For the signal extraction method, $\mathbf{V}_\zeta$ is required in the benchmarking formulae, and in this section as we assume that $\mathbf{V}_\zeta$ is known; hence a specified model is also needed to calculate $\mathbf{V}_\zeta$. Here, we let $\zeta(t)$ follow the seasonal MA model as follows.

$$\zeta(t) = (1 - \theta_\eta B)(1 - \Theta_\eta B^{12})a_\eta(t) \tag{2.1}$$

where $a_\eta(t)$ is a white noise with mean zero and variance σ_η^2. The autocovariance function $v_\zeta(k)$ of $\zeta(t)$ is as follows:

$$\begin{aligned}
v_\zeta(0) &= \sigma_\eta^2(1 + \theta_\eta^2 + \Theta_\eta^2 + \theta_\eta^2\Theta_\eta^2) && (2.2)\\
v_\zeta(1) &= -\sigma_\eta^2\theta_\eta(1 + \Theta_\eta^2) && (2.3)\\
v_\zeta(11) &= -\sigma_\eta^2\theta_\eta\Theta_\eta && (2.4)\\
v_\zeta(12) &= \sigma_\eta^2\Theta_\eta(1 + \theta_\eta^2) && (2.5)\\
v_\zeta(13) &= v_\zeta(11) && (2.6)\\
v_\zeta(k) &= 0, \; k \neq 0, 1, 11, 12, 13. && (2.7)
\end{aligned}$$

We take $\theta_\eta = 0.8$ and $\Theta_\eta = 0.6$. In fact, models for $\zeta(t)$ and the specification of their coefficients usually have no significant effect on the simulation conclusion. We only report the simulation results with $\zeta(t)$ defined by (2.1) with the abovementioned specification. In fact, we also worked on some other models, such as $\zeta(t)$ following the seasonal AR model $(1 - \phi_\eta B)(1 - \Phi_\eta B^{12})\zeta(t) = a_\eta(t)$ with $\phi_\eta = 0.8$ and $\Phi_\eta = 0.6$. The results are very similar.

However, the ratio σ_η^2/σ^2 does have a huge impact on the results. This ratio may represent the signal-to-noise ratio (S/N). Here "signal" means the stochastic variation in $\eta(t)$. It was pointed out by several authors, e.g., Trabelsi and Hillmer (1990), and Chen, Cholette, and Dagum (1997), that if the S/N is very high, benchmarking via signal extraction leads to almost the same results as those from the regression method. The lower the S/N is, the more the reduction of the error of the benchmarking prediction via signal extraction is. Thus, our investigation about the effect of mis-specification and estimation of the survey error model is combined with different choices of S/N. We take $\sigma_\eta = 3,\ 1$, or $1/3$ which represent situations of high, medium, and low signal-to-noise ratios (S/N) respectively. Note that we always let $\sigma = 1$, then $\sigma_\eta^2/\sigma^2 = 9,\ 1$ and $1/9$ respectively.

For each set of parameters, we generate data $e(t),\ t = 1, \ldots, 132$ (11 years), according to (1.3) and $\eta(t)$ according to (1.14) and (2.1). Then $y(t),\ t = 1, \ldots, 132$ and $z(T),\ T = 1, \ldots, 10$, are obtained as (1.1) and (1.2). Here we assume that year 11 has no benchmark. This is a very common situation in practice as the report for the last year may be unavailable because of delay. For each set of parameters, we repeat the data generation $10,000$ times. The data of the j^{th} replication are denoted by $\eta^{(j)}(t)$, and the corresponding benchmarking predictions (BMPs) are denoted by $\hat{\eta}^{(j)}(t), t = 1, 2, \ldots, 132$. The performance of a benchmarking method is measured by the root-mean-squared error ($RMSE$) for month t and for year T which are respectively defined as

$$\{\frac{1}{10,000}\sum_{j=1}^{10,000}[\hat{\eta}^{(j)}(t) - \eta^{(j)}(t)]^2\}^{1/2},\ t = 1, \ldots, 132, \tag{2.8}$$

and

$$\{\frac{1}{120,000}\sum_{t\in T}\sum_{j=1}^{10,000}[\hat{\eta}^{(j)}(t) - \eta^{(j)}(t)]^2\}^{1/2},\ T = 1, \ldots, 11. \tag{2.9}$$

Usually, estimates of $v_e(0)$ are given (say, from publications of statistical agencies). For investigating the impact of parameter mis-specification on the BMPs, we assume $v_e(0)$ is known. [Note that ϕ and σ^2 in (1.3) are still unknown, but they have the relationship $\sigma^2 = v_e(0)(1 - \phi^2)$.] Thus, if ϕ is mis-specified as $\tilde{\phi}$, then $v_e(k)$ are mis-specified as $\tilde{v}_e(k) = v_e(0)\tilde{\phi}^k,\ k = 0, 1, 2, \ldots$ and the $\tilde{v}_e(k)$ are used to form $\mathbf{V_e}$ in the benchmarking formulae. Some different values of $\tilde{\phi}$ in the range of $[0, 0.99]$ are tried. "$\tilde{\phi} = \phi$" (consequently, $\tilde{\sigma}^2 = 1$) means the "correct model" for the survey error. Note that in (1.11), the formula of the regression method, $\mathbf{V_e}$ can be replaced by $\mathbf{\Phi}$; then only $\tilde{\phi}$ is used to replace ϕ in $\mathbf{\Phi}$, and hence $v_e(0)$ becomes irrelevant. Also note that the results of the regression method with $\tilde{\phi} = 0.99$ can be regarded as the results of the modified Denton method.

Table 2.1 *RMSE* of BMP for different methods, true $\phi = 0.5$

		$\tilde{\phi}$							
$y/y.m$	Method	0	0.3	0.5	0.7	0.9	0.99	$\hat{\phi}$	Denton
1	Reg(1)	1.02	1.01	1.01	1.01	1.02	1.03	1.01	1.02
	SE(3)	0.96	0.94	0.94	0.94	0.97	1.02	0.94	
	SE(1)	0.72	0.70	0.69	0.70	0.78	0.96	0.69	
	SE(1/3)	0.42	0.42	0.41	0.42	0.47	0.75	0.41	
6	Reg(1)	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
	SE(3)	0.94	0.92	0.91	0.92	0.95	1.00	0.91	
	SE(1)	0.67	0.64	0.64	0.65	0.74	0.94	0.64	
	SE(1/3)	0.38	0.38	0.38	0.38	0.43	0.72	0.38	
10	Reg(1)	1.02	1.01	1.01	1.01	1.02	1.03	1.01	1.03
	SE(3)	0.95	0.93	0.92	0.93	0.96	1.01	0.93	
	SE(1)	0.69	0.67	0.66	0.67	0.76	0.97	0.66	
	SE(1/3)	0.41	0.40	0.40	0.40	0.45	0.73	0.40	
11	Reg(1)	1.16	1.15	1.15	1.15	1.18	1.30	1.16	1.33
	SE(3)	1.10	1.08	1.07	1.08	1.14	1.27	1.07	
	SE(1)	0.85	0.82	0.81	0.82	0.91	1.13	0.81	
	SE(1/3)	0.48	0.47	0.46	0.47	0.52	0.80	0.46	
1.1	Reg(1)	1.08	1.07	1.07	1.07	1.08	1.11	1.07	1.09
	SE(3)	1.01	0.99	0.99	1.00	1.02	1.10	0.99	
	SE(1)	0.75	0.73	0.73	0.74	0.79	1.01	0.73	
	SE(1/3)	0.43	0.42	0.42	0.43	0.46	0.72	0.42	
10.12	Reg(1)	1.08	1.07	1.07	1.07	1.08	1.11	1.07	1.12
	SE(3)	1.01	0.97	0.97	0.97	1.00	1.10	0.97	
	SE(1)	0.72	0.69	0.68	0.69	0.77	0.98	0.68	
	SE(1/3)	0.41	0.41	0.40	0.41	0.46	0.72	0.40	
11.12	Reg(1)	1.16	1.16	1.16	1.16	1.17	1.31	1.16	1.36
	SE(3)	1.10	1.08	1.07	1.08	1.14	1.28	1.10	
	SE(1)	0.89	0.86	0.86	0.87	0.96	1.15	0.86	
	SE(1/3)	0.51	0.49	0.49	0.50	0.53	0.78	0.49	

Note: $v_e(0)^{1/2} = 1.16$.

Table 2.2 $RMSE$ of BMP for different methods, true $\phi = 0.9$

$y/y.m$	Method	$\tilde{\phi}$ 0	0.5	0.7	0.9	0.95	0.99	$\hat{\phi}$	Denton
1	Reg(1)	1.27	1.23	1.22	1.21	1.21	1.23	1.22	1.30
	SE(3)	1.32	1.16	1.08	1.01	1.03	1.15	1.06	
	SE(1)	0.84	0.78	0.74	0.68	0.71	0.90	0.71	
	SE(1/3)	0.44	0.43	0.43	0.42	0.43	0.54	0.42	
6	Reg(1)	1.26	1.19	1.16	1.15	1.15	1.15	1.15	1.15
	SE(3)	1.29	1.11	1.02	0.93	0.96	1.08	0.99	
	SE(1)	0.74	0.72	0.66	0.62	0.64	0.83	0.64	
	SE(1/3)	0.40	0.40	0.39	0.39	0.39	0.49	0.39	
10	Reg(1)	1.27	1.23	1.21	1.20	1.21	1.22	1.21	1.22
	SE(3)	1.30	1.14	1.05	0.97	1.00	1.13	0.99	
	SE(1)	0.81	0.75	0.70	0.65	0.67	0.85	0.68	
	SE(1/3)	0.43	0.42	0.41	0.41	0.41	0.52	0.41	
11	Reg(1)	2.29	2.23	2.18	2.08	2.13	2.35	2.14	2.45
	SE(3)	2.16	1.88	1.74	1.62	1.67	2.08	1.72	
	SE(1)	1.32	1.06	0.97	0.92	0.94	1.25	0.96	
	SE(1/3)	0.53	0.49	0.48	0.47	0.48	0.60	0.48	
1.1	Reg(1)	1.61	1.59	1.59	1.55	1.57	1.62	1.58	1.92
	SE(3)	1.47	1.37	1.32	1.26	1.30	1.50	1.32	
	SE(1)	0.88	0.85	0.82	0.78	0.81	1.08	0.80	
	SE(1/3)	0.46	0.45	0.44	0.43	0.44	0.56	0.44	
10.12	Reg(1)	1.61	1.59	1.58	1.54	1.56	1.60	1.57	1.63
	SE(3)	1.46	1.30	1.22	1.14	1.18	1.43	1.20	
	SE(1)	0.87	0.79	0.74	0.69	0.71	0.94	0.72	
	SE(1/3)	0.46	0.45	0.44	0.43	0.44	0.57	0.43	
11.12	Reg(1)	2.31	2.31	2.30	2.27	2.34	2.68	2.30	2.86
	SE(3)	2.28	2.11	1.99	1.87	1.93	2.39	1.97	
	SE(1)	1.52	1.23	1.14	1.09	1.11	1.45	1.13	
	SE(1/3)	0.60	0.54	0.53	0.52	0.53	0.64	0.53	

Note: $v_e(0)^{1/2} = 2.29$.

For investigating the impact of the estimate $(\hat{\phi}, \hat{\sigma}^2)$ obtained by the survey-error-modelling procedure of Chen and Wu (2000, 2001) on the BMPs, we make the situation more practical. Since in practice the provided values of $v_e(0)$ always have errors, thus in the simulation we generate them following a distribution with the true value of $v_e(0)$ as its mean. In each replication, such a generated value is used for obtaining $(\hat{\phi}, \hat{\sigma}^2)$ by the procedure. Then, we use $\hat{v}_e(k) = \hat{\phi}^k \hat{\sigma}^2/(1 - \hat{\phi}^2)$ [see (1.12)] to replace $v_e(k)$ in $\mathbf{V}_e$ in the benchmarking formulae. That means, we also revise the provided estimates of $v_e(0)$ by $\hat{v}_e(0) = \hat{\sigma}^2/(1-\hat{\phi}^2)$. Again, for (1.11), the formula of the regression method, only $\hat{\phi}$ is needed and $\hat{v}_e(0)$ becomes irrelevant.

Tables 2.1 and 2.2 list the $RMSE$ of benchmarking prediction (BMP) by various benchmarking methods when the true $\phi = 0.5$ and 0.9. For $\phi = 0.5$, we try $\tilde{\phi} = 0,\ 0.3,\ 0.5,\ 0.7,\ 0.9$ and 0.99; for $\phi = 0.9$, we try $\tilde{\phi} = 0,\ 0.5,\ 0.7,\ 0.9,\ 0.95$ and 0.99. The columns $\hat{\phi}$ are the results of using $(\hat{\phi}, \hat{\sigma}^2)$ to form $\mathbf{V_e}$. Note that only the $RMSEs$ for a middle year (year 6) and ending years or months are listed. The notation $y/y.m$ in column 1 represents either the year or the year.month. The notations Reg(q) and SE(q), $q = 3, 1, 1/3$, represent respectively via regression and signal extraction when $\sigma_\eta = 3, 1, 1/3$. The S/N changes with σ_η as we always put $\sigma^2 = 1$.

The values in the rows of Reg(1) are from $\sigma_\eta = 1$. For the regression method, as we have pointed out, $\mathbf{V_e}$ in (1.11) can be replaced by $\mathbf{\Phi}$. From (1.4), we may write (1.11) as $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta} + \mathbf{e} + \mathbf{\Phi}\mathbf{L}^{'}(\mathbf{L}\mathbf{\Phi}\mathbf{L}^{'})^{-1}\mathbf{L}\mathbf{e}$. Hence $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}$ depends only on ϕ and $\mathbf{e}$. Thus the $RMSEs$ are the same for all choices of σ_η as long as ϕ and $\mathbf{e}$ are the same. This is confirmed by our simulation: the rows of Reg(3) and Reg(1/3) under the same $\tilde{\phi}$ are the same as that of Reg(1). On the other hand, in the column $\hat{\phi}$, $\mathbf{V_e}$ calculated from $(\hat{\phi}, \hat{\sigma}^2)$ may change when S/N changes, since $(\hat{\phi}, \hat{\sigma}^2)$ is an estimate of (ϕ, σ^2), which may depend on $\boldsymbol{\eta}$. However our simulation result shows that the differences are negligible. Hence the rows of Reg(3) and Reg(1/3) are omitted.

Since the original Denton method has nothing to do with either $\tilde{\phi}$ or $\hat{\phi}$, its $RMSEs$ are listed in the last column and put in the rows of Reg(1) for clear comparison with the regression method, especially with the modified Denton method (approximated by the regression method with $\tilde{\phi} = 0.99$)

In the following, $RMSE \sim v_e(0)^{1/2}$ represents the situation that the $RMSE$ and $v_e(0)^{1/2}$ are very close so that the BMP makes little or no improvement over the original data $y(t)$. $RMSE < v_e(0)^{1/2}$ means that the BMP is helpful and $RMSE \ll v_e(0)^{1/2}$ means that it is very helpful. $RMSE > v_e(0)^{1/2}$ means that the BMP is harmful (if we carry out benchmarking in such a way and in such a case) and $RMSE \gg v_e(0)^{1/2}$ means that it is very harmful. From Table 2.1 and 2.2, we observe the following: ($\tilde{\phi} \downarrow 0$ means $\tilde{\phi}$ decreasing and it is close to 0; $\tilde{\phi} \uparrow 1$ means $\tilde{\phi}$ increasing and it is close to 1.)

- For the regression method with a specified $\tilde{\phi}$:

  1. In the years with benchmarks (years 1 to 10): For "low ϕ", $RMSE < v_e(0)^{1/2}$ holds; the change in the $RMSE$ is very small for all $\tilde{\phi} \in [0, 0.99]$. For "high ϕ", $RMSE \ll v_e(0)^{1/2}$ alway holds and the $RMSE$ slightly increases when $\tilde{\phi} \downarrow 0$.

  2. In the year without a benchmark (year 11): For "low ϕ", $RMSE \sim v_e(0)^{1/2}$ when $\tilde{\phi}$ is not large, $RMSE > v_e(0)^{1/2}$ for larger $\tilde{\phi}$ and $RMSE \gg v_e(0)^{1/2}$ when $\tilde{\phi} \uparrow 1$. For "high ϕ", $RMSE \ll v_e(0)^{1/2}$ or $RMSE < v_e(0)^{1/2}$ holds when $\tilde{\phi} = \phi$ but the $RMSE$ increases rapidly when $\tilde{\phi}$ departs from ϕ in either direction. $RMSE \gg v_e(0)^{1/2}$ may happen when $\tilde{\phi} \uparrow 1$ and the "forecasting" lag increases.

  3. To compare with the original Denton method in the years with benchmarks: For "low ϕ", the regression method is almost the same as the Denton method no matter what $\tilde{\phi}$ is. For "high ϕ", in the middle years, if $\tilde{\phi}$ is not too low, the regression method is almost the same as the original Denton method and worse if $\tilde{\phi}$ is too low; in early months of year 1, for whatever ϕ in [0,0.99], the regression method (and hence the modified Denton method) is much better than the original Denton method.

  4. To compare with the original Denton method in the year without a benchmark: In every case, it is always better than the original Denton method which is harmful.

- Using $\tilde{\phi}$ obtained by the procedure of Chen and Wu keeps the $RMSE$ at the same level as that of using the true ϕ.

According to Points 3 and 4 above, we see that the modified Denton method is superior to the original Denton method which has already been abandoned by statistical agencies.

The General Benchmarking System developed in Statistics Canada features the modified Denton method and the regression method with the default value of $\tilde{\phi} = 0.9$ which are currently used by most users as the survey-error model is usually unavailable. According to Point 1 above, both $\tilde{\phi} = 0.9$ and 0.99 (0.99 means modified Denton method) are good choices for either "low ϕ" or "high ϕ" if the benchmarks cover the whole period where one wants to predict. However as most users are interested in predicting the variable in the current year where a benchmark may not be available, then from Point 2 above, we see that the modified Denton is very harmful and should not be recommended. With the regression method, users may tentatively use the default value $\tilde{\phi} = 0.9$, however they should try their best to obtain a reasonable estimate of ϕ for further reducing the BMP error in the year without a benchmark. We also see that if the true value of ϕ is not very high (such as $\phi < 0.95$), when the lag of "forecasting" reaches 12, the regression method shows no improvement over the original data and may be worse if $\tilde{\phi}$ remains away from the true value. For that, we should turn to the signal extraction method.

- For the signal extraction method with a specified $(\tilde{\phi}, \tilde{\sigma}^2)$ $[\tilde{\sigma}^2 = v_e(0)(1 - \tilde{\phi}^2)]$:

  1. The $RMSE$ of benchmarking prediction via signal extraction depends heavily on σ_η^2/σ^2 (the S/N). Compared with the regression method, the signal extraction method always reduces the $RMSE$ and the reduction may be drastic for low S/N. A lower S/N is more effective in reducing the $RMSE$ than having a good estimate of (ϕ, σ^2). However the S/N is not under users' control.

  2. In the years with benchmarks (years 1 to 10): For "low ϕ", the $RMSE$ increases significantly only when $\tilde{\phi} \uparrow 1$ and does not vary much for other $\tilde{\phi}$. For "high ϕ", the $RMSE$ attains a minimum at $\tilde{\phi} = \phi$ and increases when $\tilde{\phi}$ departs from the true ϕ in either direction; this phenomenon becomes less obvious for low S/N.

  3. In the year without a benchmark: The abovementioned increases, for either "low ϕ" or "high ϕ", become more significant; we may have $RMSE > v_e(0)^{1/2}$ if $\phi \uparrow 1$ when the S/N is high.

4. If the S/N is not low, for years either with or without benchmarks, a reasonable estimate is useful in a case of "high ϕ".

- For the signal extraction method, using $(\hat{\phi}, \hat{\sigma}^2)$ obtained by the procedure of Chen and Wu keeps the $RMSE$ at the same level as that of using the true (ϕ, σ^2).

3. Signal extraction method with unknown autocovariance of signal

Benchmarking via signal extraction has its advantages however it requires knowledge of the autocovariance structure of the "signal" [under assumption (1.14), it is $\mathbf{V}_\zeta$]. In practice, $\mathbf{V}_\zeta$ is unknown. In this section, we carry out a simulation study for the signal extraction method only and in almost the same way as in the previous section except that the elements $v_\zeta(k)$ of $\mathbf{V}_\zeta$ are estimated by the nonparametric method proposed by Chen, Chollete, and Dagum (1997).

For estimating $v_\zeta(k)$, at first we assume the model for the survey error $e(t)$ is given. For simplicity of statement, here we specify this model as (1.3) with parameters (ϕ, σ^2), either the true values or some other values. The major steps of the method of Chen, Cholette and Dagum (1997) are as follows.

Step 1 Let

$$\nabla\nabla_{12}y(t) = w(t) = \zeta(t) + e^*(t),\ \ t = 14, \ldots, n; \tag{3.1}$$

where $e^*(t) = \nabla\nabla_{12}e(t)$. The sample autocovariances of $w(t)$ are

$$\hat{v}_w(k) = \frac{1}{n-13}\sum_{t=14}^{n-k} w(t)w(t+k),\ \ k = 0, \ldots, n-14. \tag{3.2}$$

An estimate of the spectral density of $w(t)$ is

$$\hat{f}_w(\lambda) = \frac{1}{2\pi}[\hat{v}_w(0) + 2\sum_{k=1}^{n-14}\hat{v}_w(k)\cos(k\lambda)], \tag{3.3}$$

and the values of $\hat{f}_w(\lambda)$ at $\lambda = \lambda_j = \pi j/10n,\ j = 0, 1, \ldots, 10n$, can be calculated.

Step 2 Calculate the estimate of the spectral density of $\zeta(t)$ as

$$\tilde{f}_\zeta(\lambda) = \max\{\hat{f}_w(\lambda) - f_{e^*}(\lambda), 0\} \tag{3.4}$$

at the frequencies λ_j, where $f_{e^*}(\lambda)$ is the theoretical spectral density of $e^*(t)$,

$$f_{e^*}(\lambda) = \frac{|(1-e^{i\lambda})(1-e^{12i\lambda})|^2}{|1-\phi e^{i\lambda}|^2}\frac{\sigma^2}{2\pi} = \frac{4(1-\cos\lambda)[1-\cos(12\lambda)]}{1-2\phi\cos\lambda+\phi^2}\frac{\sigma^2}{2\pi}. \tag{3.5}$$

Step 3 Calculate

$$\tilde{v}_\zeta(k) = 2\int_0^\pi \tilde{f}_\zeta(\lambda)\cos(k\lambda)d\lambda \approx \frac{2\pi\sum_{j=1}^{10n}\tilde{f}_\zeta(\lambda_j)\cos(k\lambda_j)}{10n}, \tag{3.6}$$

and finally, the estimates of $v_\zeta(k)$ are given by

$$\hat{v}_\zeta(k) = \begin{cases} \tilde{v}_\zeta(k)u(\frac{k}{M}), & k = 0, \ldots, M, \\ 0, & M < k \le n-14. \end{cases} \tag{3.7}$$

As suggested by Chen, Cholette, and Dagum (1997), M is about $n/3$ (we use $M = 40$ in this simulation); $u(x)$ is the Parzen window given by

$$u(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & |x| \le 0.5, \\ 2(1-|x|)^3, & 0.5 < |x| \le 1, \\ 0, & \text{elsewhere}. \end{cases} \tag{3.8}$$

Table 3.1 $RMSE$ of BMP using estimated $\mathbf{V}_\zeta$, true $\phi = 0.5$

		ϕ						
$y/y.m$	σ_η	0	0.3	0.5	0.7	0.9	0.99	$\hat{\phi}$
1	3	0.98	0.96	0.95	0.96	0.98	1.01	0.95
	1	0.83	0.78	0.77	0.80	0.90	1.00	0.78
	1/3	0.66	0.62	0.62	0.69	0.85	1.00	0.66
6	3	0.97	0.94	0.93	0.93	0.96	1.00	0.93
	1	0.77	0.72	0.71	0.74	0.86	0.98	0.72
	1/3	0.58	0.53	0.53	0.61	0.80	0.98	0.60
10	3	1.00	0.97	0.96	0.96	0.99	1.03	0.96
	1	0.81	0.75	0.74	0.77	0.88	1.02	0.75
	1/3	0.63	0.57	0.58	0.64	0.82	1.01	0.57
11	3	1.13	1.10	1.10	1.11	1.16	1.31	1.11
	1	0.98	0.92	0.90	0.93	1.03	1.27	0.92
	1/3	0.84	0.77	0.75	0.78	0.94	1.26	0.79
11.12	3	1.14	1.13	1.13	1.13	1.16	1.31	1.13
	1	0.99	0.96	0.95	0.97	1.06	1.27	0.97
	1/3	0.87	0.83	0.81	0.83	0.96	1.26	0.85

Note: $v_e(0)^{1/2} = 1.16$.

Table 3.2 $RMSE$ of BMP using estimated $\mathbf{V}_\zeta$, true $\phi = 0.9$

		$\tilde{\phi}$						
$y/y.m$	σ_η	0	0.5	0.7	0.9	0.95	0.99	$\hat{\phi}$
1	3	1.46	1.20	1.09	1.03	1.07	1.17	1.09
	1	1.02	0.85	0.77	0.80	0.92	1.13	0.77
	1/3	0.71	0.52	0.49	0.66	0.85	1.12	0.61
6	3	1.52	1.23	1.09	0.97	1.00	1.10	1.07
	1	0.90	0.78	0.72	0.71	0.83	1.06	0.70
	1/3	0.61	0.45	0.43	0.56	0.75	1.04	0.52
10	3	1.54	1.27	1.14	1.03	1.07	1.19	1.12
	1	1.01	0.85	0.76	0.77	0.89	1.15	0.76
	1/3	0.72	0.53	0.49	0.63	0.82	1.13	0.59
11	3	2.34	2.03	1.87	1.76	1.87	2.29	1.87
	1	2.03	1.66	1.42	1.33	1.58	2.22	1.39
	1/3	1.88	1.46	1.19	1.16	1.48	2.20	1.20
11.12	3	2.49	2.34	2.19	2.05	2.15	2.63	2.17
	1	2.38	2.11	1.84	1.66	1.88	2.56	1.77
	1/3	2.28	1.99	1.66	1.50	1.80	2.55	1.59

Note: $v_e(0)^{1/2} = 2.29$.

When we carry out these steps in the simulation, we let (ϕ, σ^2) in (3.5) (in Step 2) take the trial values $(\tilde{\phi}, \tilde{\sigma}^2)$ $[\tilde{\sigma}^2 = v_e(0)(1-\tilde{\phi}^2)]$ no matter they are true or false, or take the estimate $(\hat{\phi}, \hat{\sigma}^2)$ obtained by the procedure of Chen and Wu (2000, 2001). Usually the estimates $\hat{v}_\zeta(k)$ of $v_\zeta(k)$ are the best when $(\tilde{\phi}, \tilde{\sigma}^2) = (\phi, 1)$, but it is not necessarily always like that (see discussion point 3 below and Appendix B). Thus the $RMSE$ of the BMP may not necessarily attain its minimum at $(\tilde{\phi}, \tilde{\sigma}^2) = (\phi, 1)$.

Table 3.1 and 3.2 list the $RMSEs$ of the BMPs via signal extraction for some years and the last month of year 11 (the end year without benchmark) when $\hat{v}_\zeta(k)$ is used in the entries of $\mathbf{V}_\zeta$. For the parameters of the survey-error model specified as $(\tilde{\phi}, \tilde{\sigma}^2)$ $[\tilde{\sigma}^2 = v_e(0)(1-\tilde{\phi}^2)]$, we observe the following:

1. For the same $\tilde{\phi}$ and σ_η, the $RMSE$ increases from the replacement of $v_\zeta(k)$ (the true value) by $\hat{v}_\zeta(k)$; the percentage increase of the $RMSE$ from this replacement is more when the S/N decreases (see Table 2.1 and 2.2 for comparison). However, to

compare these $RMSE$ from using $\hat{v}_\zeta(k)$ with those from the regression method, we roughly see that: When the S/N is high (say $\sigma_\eta = 3$), it is smaller for $\tilde{\phi} \in [0, 0.99]$ in a "low ϕ" case, and for $\tilde{\phi} \in (0.5, 0.99]$ in a "high ϕ" case; when the S/N is low (say $\sigma_\eta = 1/3$), except $\tilde{\phi} \downarrow 0$ and $\tilde{\phi} \uparrow 1$, for all other $\tilde{\phi}$ it is *much* smaller either in a "low ϕ" case or in a "high ϕ" case.

2. When $\tilde{\phi}$ increases from the true ϕ: The $RMSE$ always increases; the lower the S/N, the faster the increase. When $\phi \uparrow 1$, the $RMSE$ converges to a limit which does not depend on S/N. Hence, using a $\tilde{\phi}$ very close to 1 loses more advantage in a low S/N case. In a year with benchmark, this limit is smaller than $v_e(0)^{1/2}$ (more evident for "high ϕ"). In the year without benchmark, this limit may be much larger than $v_e(0)^{1/2}$.

3. When $\tilde{\phi}$ decreases from the true ϕ: For "low ϕ", the $RMSE$ increases slightly. For "high ϕ" and high S/N, the $RMSE$ increases and significantly when $\tilde{\phi}$ is close and outside the lower end of the "high ϕ" range. For "high ϕ" and medium or low S/N, the $RMSE$ maybe decreases at the beginning and then increases: the lower the S/N, the lower the $\tilde{\phi}$ where the $RMSE$ reaches the minimum. The reason is explained in Appendix B. However, in the year without benchmark the $RMSE$ reaches the minimum at the true ϕ in any case.

4. In the year without benchmark, for "high ϕ", the value of $\tilde{\phi}$ is very sensitive to the $RMSE$. A small deviation of $\tilde{\phi}$ from ϕ, either less or more, may lead to a large increase of the $RMSE$. This is the case where a good estimate of (ϕ, σ^2) is the most crucial.

From above discussion, we see that, in practical situations ($\mathbf{V}_\zeta$ is unknown), a good value of $(\tilde{\phi}, \tilde{\sigma^2})$ is much more important for the signal extraction method than in the situation where $\mathbf{V}_\zeta$ is known (unreal). The reason is clear: $(\tilde{\phi}, \tilde{\sigma^2})$ is not only in the BMP formulae; it also plays a role in estimating $\mathbf{V}_\zeta$. From Table 3.1 and 3.2 we observe that, using $(\hat{\phi}, \hat{\sigma}^2)$ obtained by the procedure of Chen and Wu, the $RMSE$ is usually at the same level as that from using the true (ϕ, σ^2). For "high ϕ", when the S/N is not high (the cases where a lower $\tilde{\phi}$ may lead to a smaller $RMSE$), $(\hat{\phi}, \hat{\sigma}^2)$ may even lead to a smaller $RMSE$ than that of using the true (ϕ, σ^2). These demonstrate that the survey-error-modelling procedure suggested by Chen and Wu (2000, 2001) combined with the nonparametric method of estimating $\mathbf{V}_\zeta$ proposed by Chen, Cholette, and Dagum (1997) may provide very good results of BMP in practice. The advantage of this strategy is that all statistical information is "cooked" from the data: monthly survey $y(t)$ and annual benchmarks $z(T)$.

Appendix A. Survey error model and a modelling procedure

Estimates of $v_e(0)$, the variance of survey error $e(t)$, are usually obtained in the survey process and published by statistical institutions; but estimates of the autocorrelation $v_e(k)/v_e(0)$, or equivalently, the model for $e(t)$, are rarely given.

If values of individual units (or other elementary estimates such as values for rotating panels) obtained in a survey process are available, the methods of estimating the autocorrelation or modelling the survey error are referred to as *primary analyses* of survey error (e.g. Scott, Smith and Jones, 1977; Pfeffermann, 1991). Unfortunately, in most practical situations, these values are not available due to various reasons such as confidentiality, inadequacy of data identification, and the problem of accessibility. In such cases, the methods of modelling survey error (or estimating autocorrelation of survey error) have to rely on the aggregated data, $y(t)$, in the published form. These methods are referred to as *secondary analysis* of survey error. For most of these methods, observations $y(t)$ are fitted with an integrated autoregressive moving-average (ARIMA) model which can be decomposed into two component models for the variable $\eta(t)$ and the survey error $e(t)$. Unfortunately, the conditions for such a decomposition are very restrictive which, combined with the lack of auxiliary information, enforce 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 the same.

Chen and Wu (2000, 2001) proposed a method of modelling survey error by using aggregated survey data as well as benchmarks. As more information is used in the procedure than in the secondary analysis, we refer to this kind of approach as *extended secondary analysis* of survey error. Unlike assuming very restrictive models for the variable, Chen and Wu assume only that the variable of interest $\eta(t)$ follows the DS model (1.14). Assuming that the survey error follows (1.3), the method of modelling survey error simply involves the estimation of ϕ and σ^2 using $\{y(t),\ t = 1, \cdots, n\}$ and $\{z(T),\ T = 1, \cdots, N\}$.

The basic idea is as follows. From $\{y(t)\}$ and $\{z(T)\}$, calculate

$$y^*(t) = \nabla\nabla_{12}\, y(t),\ t = 13, \cdots, n, \tag{A.1}$$

and

$$\varepsilon(T) = \sum_{t=0}^{11} y(12T - t) - z(T) = \sum_{t=0}^{11} e(12T - t),\ T = 1, \cdots, N. \tag{A.2}$$

Define the sample covariance between $y^*(k)$ and $\varepsilon(1)$ as

$$\hat{v}_{y^*\varepsilon}(k) = \frac{1}{2(N-1)} \left\{ \sum_{T=1}^{N-1} y^*(12(T-1)+k)\varepsilon(T) + \sum_{T=2}^{N} y^*(12(T+1)-(k-2))\varepsilon(T) \right\}, \tag{A.3}$$

for $14 \le k < 25$ ($k = 25$ is slightly different, see Chen and Wu, 2001). Assume the series of survey error $e(t)$ and the series of target variable $\eta(t)$ are mutually uncorrelated and let $e^*(t) = \nabla\nabla_{12}\, e(t)$; then $\hat{v}_{y^*\varepsilon}(k)$ are estimates of $v_{e^*\varepsilon}(k)$, the covariance between $e^*(k)$ and $\varepsilon(1)$ which can be expressed by the second term in the braces of $(A.4)$ below. We may get an estimate of (ϕ, σ^2), denoted by (ϕ_1, σ_1^2), by minimizing (nonlinear LS estimation)

$$S(\phi, \sigma^2) = \sum_{k=14}^{25} \{\hat{v}_{y^*\varepsilon}(k) - \rho(k, \phi)\sigma^2/(1-\phi^2)\}^2, \tag{A.4}$$

where

$$\rho(k, \phi) = -2\phi^{k-13} + \phi^{25-k} + \phi^{k-1}, \quad k = 14, 15, \cdots. \tag{A.5}$$

However this minimization gives good estimates only when N, the number of benchmarks, is large; say, a hundred or more. Unfortunately, in reality, only a few benchmarks are available. To overcome the difficulty, Chen and Wu (2000, 2001) studied the behavior of the target function $\rho(k, \phi)$ and captured some useful features of its sample counterpart in order to improve the estimates significantly. Thus, some technical strategies are involved in the practical implementation of the method. For convenience, an outline of the procedure is presented in the following. For more details, see Chen and Wu (2001). Note that in the minimization procedures in Steps 1 and 5, the areas for searching optimal values of ϕ and/or σ^2 are some neighbourhoods of (ϕ, σ^2) satisfying the relationship $\sigma^2 = v_e(0)/(1-\phi^2)$; where $v_e(0)$ is the variance of the survey error and its estimate is usually published by statistical agencies or provided by survey experts.

Step 1 As described in the above, a tentative estimate (ϕ_1, σ_1^2) is obtained by minimizing (A.4).

Step 2 Use some auxiliary features (statistics) and ϕ_1 to obtain a reasonable estimate ϕ_2 and to test if the true ϕ is "low" or "high" (if there is no prior knowledge about ϕ).

Step 3 Trim $\hat{\rho}(k) = \hat{v}_{y^*\epsilon}(k)/\hat{v}_e(0)$ to $\rho^*(k)$ (outlier adjustment) according to "low ϕ" or "high ϕ", where $\hat{v}_{y^*\epsilon}(k)$ is given by (1.9). Obtain ϕ_3 by minimizing

$$S_R^*(\phi) = \sum_{k=14}^{20} [R^*(k) - R(k,\phi)]^2, \quad R^*(k) = \rho^*(k)/G(\phi_2). \tag{A.6}$$

where

$$R(k,\phi) = \rho(k,\phi)/G(\phi), \tag{A.7}$$

$\rho(k,\phi)$ is given by (A.5) and

$$G(\phi) = \{\frac{1}{11}\sum_{k=15}^{25} [\rho(k,\phi) - \rho(k-1,\phi)]^2\}^{1/2}. \tag{A.8}$$

Step 4 For further improving ϕ_3, re-define $\hat{\rho}(k)$ as

$$\hat{\rho}(k) = \hat{v}_{e^*\epsilon}(k)/[\hat{v}_e(0)(1-\phi_3^2)/(1-\phi_2^2)]. \tag{A.9}$$

and then get $\rho^*(k)$ as in Step 3. Redo the minimization (A.6) to obtain an optimal value ϕ_4 which is the final estimate of ϕ.

Step 5 The final estimate σ_4^2 of σ^2 based on ϕ_4 is obtained by minimizing (A.4) in a certain region of σ^2 with ϕ replaced by ϕ_4 in (A.4).

Appendix B. The minimum of the $RMSE$ may not be at $\tilde{\phi} = \phi$

For benchmarking via signal extraction, when $v_e(k)$ are "known" (true, false or estimated), the key step for implementing benchmarking via signal extraction [(1.16) through (1.20)] in practice, is estimating $v_\zeta(k)$.

When S/N is low, $\zeta(t)$ is a weaker component in $w(t) = \zeta(t) + e^*(t)$ to make a contribution to $\hat{f}_w(\lambda)$, the true spectrum of $e^*(t)$ does not necessarily lead to the best estimate of $f_\zeta(\lambda)$ [see (3.4)], and hence the best $\hat{v}_\zeta(k)$. Then the $RMSE$ of the benchmarking prediction may not necessarily attain its minimum at the true value of (ϕ, σ^2). In Table B1, we list some means (each is obtained from 1,000 replications) of $\hat{v}_\zeta(k)$ (only the results for $k = 0$ and 1 are listed). The formulae for the true values of $v_\zeta(k)$ are given by (2.2) through (2.7).

From this table, we see that, when $\sigma_\eta = 3$ and 1, i.e. the situations of high and medium S/N, the mean of $\hat{v}_\zeta(k)$ $(k = 0, 1)$ is the closest to the true values of $v_\zeta(k)$ at $\tilde{\phi} = 0.9$. However when $\sigma_\eta = 1/3$, the mean of $\hat{v}_\zeta(0)$ is the closest to (sometimes the same as) the true value of $v_\zeta(0)$ at $\tilde{\phi} = 0.7$. Even $\tilde{\phi} = 0.5$ and 0 is better than $\tilde{\phi} = 0.9$ for $k = 0$; although for $k = 1$, $\tilde{\phi} = 0.9$ gives the best $\hat{v}_\zeta(1)$.

Table B1 Some means of the estimate of $v_\zeta(k)$, true $\phi = 0.9$

		$\tilde{\phi}$						
k	σ_η	0	0.5	0.7	0.9	0.95	0.99	true $v_\zeta(k)$
0	3	9.07	14.25	17.08	20.40	21.34	22.14	20.07
	1	0.354	0.542	1.145	2.750	3.463	4.184	2.230
	1/3	0.236	0.148	0.225	0.986	1.531	2.184	0.225
1	3	-3.37	-7.25	-8.64	-9.59	-9.71	-9.75	-9.79
	1	0.226	-0.089	-0.450	-0.994	-1.105	-1.149	-1.088
	1/3	0.206	0.104	0.040	-0.121	-0.170	-0.192	-0.121

Acknowledgment

Our sincere thanks are extended to Benoit Quenneville, Pierre Cholette, Siu Hung Chung, and John Higginson, for stimulating discussions and valuable comments, also for help with the editing. This research was partially supported by the Hong Kong RGC Earmark Grant CUHK4314/02H.

References

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

Chen, Z. G. and Wu, K. H. (2000). Survey error modelling in the presence of benchmarks. *Proceedings of the Survey Methods Section, Annual Meeting of Statistical Society of Canada,*, 93-101.

Chen, Z. G. and Wu, K. H. (2001). Survey error modelling in the presence of benchmarks. *Working Paper* no. BSMD-2001-013E, Statistics Canada.

Chollete, P. A. (1984). Adjusting sub-annual series to yearly benchmarks. *Survey Methodology*, **10**, 35-49.

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

Dagum, E. B., Cholette, P. A. and Chen, Z. G. (1998). A unified view of signal extraction benchmarking, interpolation and extrapolation of time series. *International Statistical Review*, **66**, 245-296.

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

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

Fernandez, R. B. (1981). A methodological note on the estimation of time series. *The Review of Economics and Statistics*, **63**, 471-476.

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

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.

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.

Ca005
C.2