

BENCHMARKING COMPARISON AND ERROR MODELLING

Ka Ho Wu¹ and Zhao-Guo Chen²

ABSTRACT

For a target socio-economic variable, two sources of data with different frequencies and precisions may be available. A benchmarking procedure is using the less frequent and more reliable data to adjust the more frequent and less reliable data. For advanced benchmarking methods, an error model for more frequent series is needed but usually unavailable. In this presentation, we compare three benchmarking methods, namely the Denton, regression, and signal-extraction methods. For the regression method, assuming the error series follows an AR(1) model, we compare the benchmarking results by using default models and by using an error-modelling procedure. This study is useful for developing and applying a benchmarking software.

KEY WORDS: Benchmarking; AR model; Survey-error modelling.

RÉSUMÉ

Deux sources de données de précisions différentes et avec des fréquences de collecte différentes peuvent être disponibles pour une variable socioéconomique. Une procédure d'étalonnage consiste à utiliser les données moins fréquentes et plus fiables pour corriger les données plus fréquentes et moins fiables. Pour les méthodes d'étalonnage avancées, un modèle d'erreurs pour les séries de données plus fréquentes est nécessaire mais généralement non disponible. Dans cet article, nous comparons trois méthodes d'étalonnage, à savoir la méthode de Denton, la méthode basée sur un modèle de régression et la méthode basée sur l'extraction du signal. En supposant que la série d'erreurs suit un processus autorégressif d'ordre 1, nous comparons les résultats de l'étalonnage qui utilise des modèles par défaut et un processus de modélisation des erreurs. Cette étude est utile pour mettre au point et appliquer un logiciel d'étalonnage.

MOTS CLES: L'étalonnage; modèle AR; modélisation de l'erreur d'enquête.

1. INTRODUCTION

In data process, benchmarking has long been recognized to be an important area and is crucial for statistical agencies and economic researchers. For a target socio-economic variable, two different sources of data with different levels of precision and collecting frequencies may be available. Typically, the less frequent data (e.g. annual report or census) is more reliable and considered as benchmarks. The process of using benchmarks to adjust the more frequent and less reliable data (say, the repeated monthly survey) is called benchmarking. Application examples of benchmarking monthly series, such as retail series, can be found in Dagum, Cholette and Chen (1998), and Dagum and Cholette (2006).

Suppose we have monthly observations $y(t)$,

$$y(t) = \eta(t) + e(t), \quad t = 1, \dots, n, \quad E e(t) = 0 \quad (1.1)$$

where $\eta(t)$ is the target socio-economic variable and $e(t)$ is the error (mainly we refer it to repeated survey-error as it is correlated and can be modeled) which is assumed to be stationary. Moreover, suppose we also have the annual sums $z(T)$, called benchmarks, which are obtained from a more reliable source, i.e.

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

¹Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong. email: wu@sta.cuhk.edu.hk

²Time Series and Analysis Centre, 17H-REC, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6. email:Chen.Zhao-Guo@statcan.gc.ca

where the notation $t \in T$ means that 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 observation error of benchmarks; its variance is usually much smaller than the variance of $e(t)$. In most 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 paper, we focus on binding benchmarking.

Assume that the survey error $\{e(t)\}$ follows a first order autoregressive [AR(1)] model

$$e(t) = \phi e(t-1) + \xi(t) \quad (1.3)$$

where $0 \leq \phi < 1$ and $\{\xi(t)\}$ is a white noise process with mean zero and variance σ^2 . The autocovariance function of $\{e(t)\}$ is denoted as $v_e(k) = \phi^k \sigma^2 / (1 - \phi^2)$, $k = 0, 1, 2, \dots$. As pointed out by Scott, Smith and Jones (1977) and Chen and Wu (2000), this assumption is reasonable in many practical applications. Unlike assuming some restrictive models for the variable, we assume the variable of interest $\eta(t)$ follows a difference stationary (DS) model without any further specifications

$$\nabla \nabla_{12} \eta(t) = \zeta(t) \quad (1.4)$$

where $\{\zeta(t)\}$ is a stationary, mean zero and possibly over-differenced time series. $\nabla = 1 - B$, $\nabla_{12} = 1 - B^{12}$, and B is the backshift operator defined by $B^k \eta(t) = \eta(t - k)$. The DS model is a very general nonstationary model which can fit many real series very well and is widely used by many authors.

In the binding case, (1.1) and (1.2) can be jointly written in the vector form:

$$\begin{cases} \mathbf{y} &= \boldsymbol{\eta} + \mathbf{e} \\ \mathbf{z} &= \mathbf{L}\boldsymbol{\eta}, \end{cases} \quad (1.5)$$

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

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

If $n > 12N$, then some columns of 0's corresponding to the months not covered by benchmarks are added to the above \mathbf{L} . In the non-binding case, an error term should be added to the right-hand side of the second equation in (1.5).

2. BENCHMARKING METHODS

In this section, we provide a brief review of three benchmarking methods, namely the Denton method, the regression method and the signal extraction method, and an example for comparing these methods. More details and further discussions can be found in Chen and Wu (2006).

Denton (1971) method is widely used by statistical agencies for benchmarking time series. With Denton method, by minimizing a penalty function, the movement of the benchmarking prediction $\hat{\boldsymbol{\eta}} = (\hat{\eta}(1), \dots, \hat{\eta}(n))'$ is kept to the closest to that of the original series $\mathbf{y} = (y(1), \dots, y(n))'$. The Denton method is a numerical approach, easy to apply and no statistical information of the error series of target series is required.

The benchmarking method introduced by Cholette and Dagum (1994) is based on a regression model. They took into account of the presence of autocorrelated error in the more frequent data. By assuming $\eta(t)$ are fixed values in model (1.5), (1.5) becomes a regression model with $n + N$ observations and n parameters. In order to minimize the variance of the estimation error, $Var(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$, the generalized least square (GLS) estimation leads to the solution (see p.367-p.368, Cholette and Dagum, 1994),

$$\hat{\boldsymbol{\eta}} = \mathbf{y} + \mathbf{V}_e \mathbf{L}' (\mathbf{L} \mathbf{V}_e \mathbf{L}')^{-1} (\mathbf{z} - \mathbf{L} \mathbf{y}) \quad (2.1)$$

where \mathbf{V}_e is the covariance matrix of \mathbf{e} . For implementing the procedure, \mathbf{V}_e is required.

By regarding the variable of interest, $\eta(t)$, as a stochastic signal following an ARIMA model, Hillmer and Trabelsi (1987) proposed a benchmarking method via signal extraction. In the first step, by using the statistical information (model or autocovariance) of $\{e(t)\}$ and $\{\eta(t)\}$, an extracted signal, $\hat{\boldsymbol{\eta}}_0$, and the covariance matrix of its prediction

error \mathbf{V}_0 are obtained. In the second step, benchmarks are used to adjust $\hat{\boldsymbol{\eta}}_0$ to obtain the final benchmarking prediction. For implementing the signal extraction method, both covariance matrices \mathbf{V}_e and \mathbf{V}_η are required.

The format of benchmarking formula for the Denton method or the signal extraction method is the same as (2.1). With Denton method, \mathbf{V}_e in (2.1) is replaced by an $n \times n$ matrix, denoted by \mathbf{A}^{-1} , where its (i, j) entry is $Min(i, j)$. With the signal extraction method, \mathbf{V}_e in (2.1) is replaced by \mathbf{V}_0 and \mathbf{y} is replaced by $\hat{\boldsymbol{\eta}}_0$. It is conceivable that, the minimum mean squared error (note that $\hat{\boldsymbol{\eta}}$ is an unbiased prediction of $\boldsymbol{\eta}$ either via the regression or the signal extraction method) of the benchmarking prediction of $\eta(t)$ based on the signal extraction method should be better than that of the regression method as the benchmarks are used to adjust the preliminary error-reduced $\hat{\boldsymbol{\eta}}_0$ rather than the original \mathbf{y} .

To provide a simulation study for comparing the methods, we assume the survey error $\{e(t)\}$ follows an AR(1) model as (1.3), where $Var(\xi(t)) = \sigma^2 = 1$ with $\phi = 0.9$ which gives $\{v_e(0)\}^{1/2} = 1/(1 - \phi^2)^{1/2} = 2.29$. Then we assume the signal $\eta(t)$ follows the DS model (1.4), with $\zeta(t)$ follows a seasonal MA model $\zeta(t) = (1 - 0.8B)(1 - 0.6B^{12})a_\eta(t)$ where the SD of $a_\eta(t)$ is $\sigma_\eta = 3$, or 1, or 1/3. $S/N = \sigma_\eta/\sigma = 3, 1, 1/3$, are the three cases corresponding to high, medium and low signal-to-noise ratio. We repeatedly generate (10,000 times) $\eta(t)$ and $e(t)$, $t = 1, \dots, 132$ (11 years) so that $y(t) = \eta(t) + e(t)$, $t = 1, \dots, 132$ and $z(T) = \sum_{t \in T} \eta(t)$, $T = 1, \dots, 10$ are created. We assume that year 11 has no benchmark due to report delay. For each of the aforementioned benchmarking methods, we obtain the (sample) root mean squared error (*RMSE*) of the benchmarking prediction for each year $T = 1, \dots, 11$. i.e. the root of the average of $\{\hat{\eta}(t) - \eta(t)\}^2$ over each month and all the replications. Table 1 lists some results.

For Denton or the regression method, the signal does not effect the *RMSE*; but the *RMSE* strongly depends on the signal for the signal extraction method (SE). Note that, using the original data $y(t)$ to predict $\eta(t)$, the prediction error is $e(t)$ itself, so that the *RMSE* is simply the standard deviation (SD) of $e(t)$, which is 2.29. Therefore, a value of *RMSE* much lower than 2.29 means the benchmarking prediction is very good in the indicated year, and a value higher than 2.29 means it is harmful. From Table 1, we can see that the SE method is always better than the regression method. The reduction of *RMSE* is particularly drastic when S/N is low [$SE(1/3)$]. The regression method does not improve the results from the Denton method in the middle period covered by benchmarks (say, year 6), but there is some gain in both end years (year 1 and 10), and the regression method is much better in the year without benchmark (year 11). In fact in this case, benchmarking using Denton method is harmful for year 11.

Table 1. *RMSE* of BMP with different methods, $\phi = 0.9$

Year	Method	RMSE	Year	Method	RMSE
1	Denton	1.30	6	Denton	1.15
	Reg	1.21		Reg	1.15
	SE(3)	1.03		SE(3)	0.96
	SE(1)	0.71		SE(1)	0.64
	SE(1/3)	0.43		SE(3)	0.39
10	Denton	1.22	11	Denton	2.45
	Reg	1.20		Reg	2.08
	SE(3)	0.97		SE(3)	1.62
	SE(1)	0.65		SE(1)	0.92
	SE(1/3)	0.41		SE(1/3)	0.47

3. BENCHMARKING VIA REGRESSION METHOD WITH A WORKING MODEL FOR DATA ERROR

The SE method is the most advanced, but it requires much more statistical information. The regression method only requires the error model, or equivalently, the covariance matrix of the error, \mathbf{V}_e . This method is more robust and easier to implement than the SE method. In this section, we focus on the regression method and discuss how to choose the error model, equivalently \mathbf{V}_e , in benchmarking process when the model is unknown. If the model [equivalently, \mathbf{V}_e] is known, benchmarking prediction $\hat{\boldsymbol{\eta}}$ can be obtained by (2.1) and the covariance matrix of the prediction error, $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}$, is given by

$$\mathbf{V}_1 = Var(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) = [\mathbf{I} - \mathbf{V}_e \mathbf{L}' (\mathbf{L} \mathbf{V}_e \mathbf{L}')^{-1} \mathbf{L}] \mathbf{V}_e [\mathbf{I} - \mathbf{V}_e \mathbf{L}' (\mathbf{L} \mathbf{V}_e \mathbf{L}')^{-1} \mathbf{L}]'. \quad (3.1)$$

In the case of unknown \mathbf{V}_e , the benchmarking procedure still works by replacing \mathbf{V}_e in (2.1) with a *working covariance-matrix* $\tilde{\mathbf{V}}_e$. If this $\tilde{\mathbf{V}}_e$ is from a model, then this model is called the *working model*. It can be a default model chosen by user, or a model obtained by a procedure. The benchmarking prediction is given by

$$\tilde{\boldsymbol{\eta}} = \mathbf{y} + \tilde{\mathbf{V}}_e \mathbf{L}' (\mathbf{L} \tilde{\mathbf{V}}_e \mathbf{L}')^{-1} (\mathbf{z} - \mathbf{L} \mathbf{y}). \quad (3.2)$$

The covariance matrix of the prediction error, $\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}$, thus becomes

$$\mathbf{V}_2 = \text{Var}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) = [\mathbf{I} - \tilde{\mathbf{V}}_e \mathbf{L}' (\mathbf{L} \tilde{\mathbf{V}}_e \mathbf{L}')^{-1} \mathbf{L}] \mathbf{V}_e [\mathbf{I} - \tilde{\mathbf{V}}_e \mathbf{L}' (\mathbf{L} \tilde{\mathbf{V}}_e \mathbf{L}')^{-1} \mathbf{L}]'. \quad (3.3)$$

Denote the diagonal elements of \mathbf{V}_i by $v_i(t)$. $\sqrt{v_2(t)}$ is the SD of the real benchmarking prediction, $\tilde{\boldsymbol{\eta}}$, at month t . $\sqrt{v_2(t)} - \sqrt{v_1(t)}$ is the increase of the SD of benchmarking prediction by not using the true covariance-matrix of $\{e(t)\}$ as working covariance-matrix. Since either $\hat{\boldsymbol{\eta}}$ or $\tilde{\boldsymbol{\eta}}$ are unbiased prediction of $\boldsymbol{\eta}$, thus $\sqrt{v_j(t)}$ is in fact equal to the *RMSE* of the corresponding prediction at month t .

However if \mathbf{V}_e , the middle factor of (3.3), is still unknown, $\sqrt{v_2(t)}$ cannot be calculated. In such a situation, one may use the same working covariance-matrix $\tilde{\mathbf{V}}_e$ to replace \mathbf{V}_e in (3.4). Hence instead of $\sqrt{v_2(t)}$, the published value is $\sqrt{v_3(t)}$ where $v_3(t)$ is the diagonal element of

$$\mathbf{V}_3 = [\mathbf{I} - \tilde{\mathbf{V}}_e \mathbf{L}' (\mathbf{L} \tilde{\mathbf{V}}_e \mathbf{L}')^{-1} \mathbf{L}] \tilde{\mathbf{V}}_e [\mathbf{I} - \tilde{\mathbf{V}}_e \mathbf{L}' (\mathbf{L} \tilde{\mathbf{V}}_e \mathbf{L}')^{-1} \mathbf{L}]'. \quad (3.4)$$

4. SOME COMPARISONS

Assuming that $\{e(t)\}$ follows an AR(1) model in (1.3). By direct calculation or simulation similar to that in Section 2, we will show how benchmarking via the regression method performs with $\tilde{\phi}$, a default value for ϕ , or with $\hat{\phi}$, the estimated value using a modelling procedure proposed by Chen and Wu (2000). For ease of comparison, we always let $v_e(0) = 1$ [the SD of $e(t)$ is also 1]. Thus, $\sigma^2 = 1 - \phi^2$. Again we assume that there are 11 years' monthly data $y(t)$ with binding benchmarks $z(T)$ from year 1 to 10 as in Section 2.

Table 2 lists the values of $\sqrt{v_1(t)}$ and Table 3 lists the values of $\sqrt{v_2(t)} - \sqrt{v_1(t)}$ using $\tilde{\phi} = 0.9$ as the working model with $\phi = 0, 0.5, 0.7, 0.8, 0.9, 0.95, 0.975, 0.995$ for some selected months. $\tilde{\phi} = 0.9$, which has a lot of advantages from the numerical point of view, is used as the default value in the software *Proc Benchmarking*. In the table, we provide the results for Month=6.1, 10.12, 11.1, 11.3, 11.6, 11.12 or equivalently $t = 61, 120, 121, 123, 126, 132$. Note that Months 6.1 and 10.12 represent respectively the middle month of the period and the end month covered by benchmarks. (Note that the result for Month 1.1 is the same as that of Month 10.12 due to symmetrical reason.) As Months 11.1, 11.3, 11.6, 11.12 are not covered by benchmarks, the benchmarking predictions in these months are considered as *preliminary* since their values may have to be adjusted a lot when a new benchmark becomes available. We only report the results for $\eta(t) = 0$. For non-zero $\eta(t)$, the result is quite the same by using a fixed default $\tilde{\phi}$, but may be slightly worse when the estimated $\hat{\phi}$ is used.

From Table 2, we see that, for benchmarking using the true ϕ , the higher the ϕ , the more the reduction [from 1, the SD of $e(t)$] of the SD of $\hat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}(t)$ is. For ϕ close to 0, benchmarking barely reduces the SD and the preliminary benchmarking prediction in year 11 is almost the same (SD=1) as using solely the original data $y(t)$.

Table 2 Some values of $\sqrt{v_1(t)}$

ϕ	Month					
	6.1	10.12	11.1	11.3	11.6	11.12
0	0.957	0.957	1.000	1.000	1.000	1.000
0.5	0.922	0.935	0.984	0.999	1.000	1.000
0.7	0.852	0.889	0.947	0.988	0.999	1.000
0.8	0.763	0.826	0.893	0.985	0.989	0.999
0.9	0.578	0.674	0.747	0.843	0.920	0.978
0.95	0.415	0.512	0.578	0.676	0.775	0.886
0.975	0.293	0.374	0.472	0.511	0.604	0.729
0.995	0.131	0.172	0.221	0.241	0.294	0.374

Table 3 provides the results of $\sqrt{v_2(t)} - \sqrt{v_1(t)}$ when $\tilde{\phi} = 0.9$ is used as the working model. As we can see, there is almost no increase in Month 6.1 for all ϕ . However, at the end of year 10 and in year 11, the increase is small when ϕ is low and the increase is large or very large when ϕ is close to 1. Note that, the increase always refers to the corresponding value in Table 2. For example, for $\phi = 0.995$ in Month 11.12, the increase 0.425 from 0.374 in Table 2 is 113.6%. It is noted that using $\tilde{\phi} = 0.9$ is usually not bad when ϕ is not very high but is bad at year 11 and the two ends of year 1 and 10 when ϕ is very high. However, it can be shown that the result is always good when the modelling procedure is used to obtain $\hat{\phi}$ for the working model, particularly when ϕ is very high.

Table 3 $\sqrt{v_2(t)} - \sqrt{v_1(t)}$, $\tilde{\phi} = 0.9$

ϕ	Month					
	6.1	10.12	11.1	11.3	11.6	11.12
0	0.017	0.003	0.034	0.022	0.012	0.003
0.5	0.008	0.010	0.031	0.043	0.029	0.009
0.7	0.004	0.011	0.024	0.037	0.033	0.013
0.8	0.001	0.007	0.012	0.020	0.021	0.011
0.9	0.000	0.000	0.000	0.000	0.000	0.000
0.95	0.000	0.008	0.013	0.023	0.031	0.034
0.975	0.000	0.028	0.046	0.077	0.111	0.135
0.995	0.000	0.093	0.141	0.225	0.320	0.425

In most situations, using $\tilde{\phi} = 0.9$ for working model is acceptable. However, it is not acceptable if it is used for obtaining $\sqrt{v_3(t)}$, as an estimate of $\sqrt{v_2(t)}$ (published as a quality measure of benchmarking prediction) since the difference $\sqrt{v_3(t)} - \sqrt{v_2(t)}$ can be very large. For example, for $\phi = 0$ in Month 6.1 ($t = 61$), $\sqrt{v_2(61)} = 0.957 + 0.017 = 0.974$ (see $\phi = 0$ and Month 6.1 in Table 2 and 3). This difference is -0.396 which gives about 40% underestimate. Even for $\phi = 0.7$ (close to 0.9) in Month 6.1, $\sqrt{v_2(61)} = 0.856$, this difference is -0.278, a 32% underestimate. From Table 4, we can see that $\sqrt{v_2(t)}$ is seriously overestimated when ϕ is much larger than 0.9. For example, for $\phi = 0.995$ in Month 6.1, $\sqrt{v_2(61)} = 0.131$, $\sqrt{v_3(61)} = 0.131 + 0.447 = 0.578$, that is a 441% overestimate.

Table 4 $\sqrt{v_3(t)} - \sqrt{v_2(t)}$ using default ($\tilde{\phi}$) and estimate ($\hat{\phi}$), $\eta(t) = 0$

ϕ	Mean/SD	$\tilde{\phi}$ or $\hat{\phi} - \phi$	Month					
			6.1	10.12	11.1	11.3	11.6	11.12
0.9		$\tilde{\phi} = 0.9$	0.000	0.000	0.000	0.000	0.000	0.000
$(\hat{\phi})$	Mean	-0.009	-0.042	-0.067	-0.084	-0.110	-0.128	-0.124
	SD	0.080	0.202	0.202	0.211	0.215	0.214	0.208
0.95		$\tilde{\phi} = 0.9$	0.163	0.154	0.156	0.144	0.113	0.058
$(\hat{\phi})$	Mean	-0.015	0.004	-0.018	-0.031	-0.055	-0.080	-0.101
	SD	0.063	0.176	0.174	0.179	0.181	0.183	0.186
0.975		$\tilde{\phi} = 0.9$	0.284	0.271	0.274	0.254	0.204	0.114
$(\hat{\phi})$	Mean	-0.005	0.002	-0.013	-0.019	-0.030	-0.045	-0.067
	SD	0.034	0.122	0.130	0.138	0.150	0.163	0.178
0.995		$\tilde{\phi} = 0.9$	0.447	0.410	0.408	0.576	0.306	0.179
$(\hat{\phi})$	Mean	0.000	-0.004	-0.006	-0.007	-0.010	-0.013	-0.019
	SD	0.004	0.143	0.058	0.066	0.080	0.095	0.116

Table 4 lists some values of $\sqrt{v_3(t)} - \sqrt{v_2(t)}$ by using $\tilde{\phi} = 0.9$ and using the estimate $\hat{\phi}$ [indicated by the column " $\tilde{\phi}$ or $\hat{\phi} - \phi$ " and rows of $(\hat{\phi})$] proposed by Chen and Wu (2000). The rows of (sample) Mean and SD are obtained from 1,000 replications of $\sqrt{v_3(t)} - \sqrt{v_2(t)}$ of using $\hat{\phi}$. Note that $\hat{\phi}$ changes from replication to replication. $\tilde{\phi} = 0.9$ is better than $\hat{\phi}$ only for ϕ in the range about [0.8, 0.95]. This table only lists those of ϕ greater than or equal to 0.9. For ϕ less than 0.8, using $\tilde{\phi} = 0.9$ is worse or much worse comparing with using $\hat{\phi}$. Also note that, except the later months in year 11 and ϕ is close to 0.9, $\sqrt{v_3(t)}$ is almost an unbiased estimate of $\sqrt{v_2(t)}$ when $\hat{\phi}$ is used.

5. CONCLUSIONS FOR ERROR MODEL IN BEHNCHMARKING VIA REGRESSION METHOD

The following conclusions may be used as references to form a guideline for choosing error model (if unknown) in benchmarking via regression method.

1. The coefficient of a working AR(1)-model almost does not affect the SD of the benchmarking-prediction error in the middle of the period covered by benchmarks; but it does affect this SD in the months at both ends of this period, and may affect more in the year without benchmark.
2. The choice of working model seriously affects the publication error of the SD of benchmarking-prediction error.
3. As the minimization of benchmarking prediction is concerned, $\tilde{\phi} = 0.9$ is a reasonable choice in general, but not good when the error of the monthly data is very highly autocorrelated. For publishing the SD of benchmarking-prediction error, using $\tilde{\phi} = 0.9$ may lead to large publication error if the true ϕ is not close to 0.9.
4. The proposed method for modelling the error series of monthly data may provide us with a good working model for benchmarking. That may reduce the SD of benchmarking-prediction error, especially may drastically reduce the publication error for this SD.

ACKNOWLEDGEMENT

This research was supported by research funding of Statistics Canada and the Hong Kong RGC Earmarked Research Grant Project No. 452607.

REFERENCES

- Chen, Z.-G. and Wu, K.H. (2000). Survey error modelling in the presence of benchmarks. *Proceedings of Survey Methods Section, Statistical Society of Canada*, 93-101.
- Chen, Z.-G. and Wu, K.H. (2006). Comparison of benchmarking methods with and without survey error modelling. *International Statistical Review*, 74, 285-304.
- Cholette, 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.
- Dagum, E.B. and Cholette, P.A. (2006). *Benchmarking, temporal distribution, and reconciliation methods for time series*. Springer, New York.
- 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.
- Hillmer, S.C. and Trabelsi, A. (1987). Benchmarking of economic time series. *Journal of American Statistical Association*, 82, 1064-1071.
- 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.