Survey Methodology, December 1993
Vol. 19, No. 2, pp. 165–172
Statistics Canada

Maximum Likelihood Estimation of Constant Multiplicative Bias Benchmarking Model with Application

IJAZ U.H. MIAN and NORMAND LANIEL[1]

ABSTRACT

The maximum likelihood estimation of a non-linear benchmarking model, proposed by Laniel and Fyfe (1989; 1990), is considered. This model takes into account the biases and sampling errors associated with the original series. Since the maximum likelihood estimators of the model parameters are not obtainable in closed forms, two iterative procedures to find the maximum likelihood estimates are discussed. The closed form expressions for the asymptotic variances and covariances of the benchmarked series, and of the fitted values are also provided. The methodology is illustrated using published Canadian retail trade data.

KEY WORDS: Autocorrelations; Bias model; Generalized least squares; Sampling errors.

1. INTRODUCTION

Benchmarking methods are very commonly used for improving sub-annual survey estimates with the help of corresponding estimates, called benchmarks, from an annual survey. The improvement generally is in terms of reductions in the biases and variances of the sub-annual estimates. For example, the monthly retail trade estimates might be improved using estimates from annual retail trade surveys. The sub-annual estimates are often biased due to coverage deficiencies in the frame. Undercoverage is caused by delay in the inclusion of new businesses and non-representation of non-employer businesses in the frame. Furthermore, the variances of the sub-annual estimates are often larger than those of the corresponding annual estimates, and the sampling covariances exist between the estimates from different time periods due to overlap of the samples. On the other hand, the annual estimates can be assumed unbiased because, in practice, their frames do not suffer much from coverage deficiencies. Detailed discussions on benchmarking can be found in Laniel and Fyfe (1989; 1990), Cholette (1987; 1988), and others.

Several procedures for benchmarking time series are available in the literature. Based on a quadratic minimization approach, Denton (1971) proposed several procedures to benchmark a single time series. Cholette (1984) proposed a modified version of Denton's order one proportional variant method where he removed the starting condition to avoid transient effects. The assumptions made by authors are very unlikely to be satisfied by most economic time series. More specifically, their models assume that the bias associated with sub-annual estimates follows a random walk and that both the sub-annual and annual data are observed without sampling errors. In general the estimates come from sample surveys and hence they are subject to sampling errors.

Hillmer and Trabelsi (1987) proposed an alternate approach to benchmarking which is based on an ARIMA model (see *e.g.*, Box and Jenkins 1976). Although this approach takes into account the sampling covariances of the sub-annual and annual estimates, the approach does not accommodate biases in the sub-annual estimates. Cholette and Dagum (1989) modified the Hillmer and Trabelsi approach by replacing the ARIMA model by an "intervention" model. This approach allows the modelling of systematic effects in the time series but still possesses the same weaknesses as found in the Hillmer and Trabelsi model (Laniel and Fyfe 1990).

In order to overcome the deficiencies mentioned above, Laniel and Fyfe (1989; 1990) proposed a non-linear benchmarking model on levels. The authors provided a complex algorithm to find the generalized least squares (GLS) estimates (and their asymptotic covariances) of the model parameters. This model takes into account the sampling covariances of the sub-annual and annual estimates, and can be used when the benchmarks come either from censuses or annual overlapping samples. This model also assumes a constant multiplicative (relative) bias associated with the sub-annual level estimates. Other constant multiplicative bias benchmarking models has been proposed by Cholette (1992) and Laniel and Mian (1991). Cholette assumes a model in which both the bias and errors are multiplicative. The author used the GLS theory to find the estimates of the model parameters after making a logarithmic transformation on the model. Laniel and Mian (1991) have provided an algorithm to find the maximum likelihood estimates of a constant multiplicative bias benchmarking model with mixed (a mixture of binding and non-binding) benchmarks. The binding benchmark here is an estimate from a census (*i.e.*, an estimate with zero variance) and the non-binding benchmark on the other hand is an estimate based on a sample. The assumption

[1] Ijaz U.H. Mian and Normand Laniel, Social Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6, Canada.

of a constant multiplicative bias will be verified in practice if the rate of frame maintenance activities is relatively stable, that is, when the proportion of frame coverage deficiencies is fairly constant over time. This assumption also implies that the covered and uncovered businesses in the frame possesses the same average period-to-period ratios with respect to the variable of interest. The nature of bias associated with sub-annual estimates may vary from one time series to another. Cholette and Dagum (1991) have proposed a benchmarking method which assumes a constant additive bias associated with the sub-annual estimates.

The purpose of this paper is to consider the maximum likelihood (ML) estimation of the parameters of Laniel and Fyfe's model and the results are based on the report of Mian and Laniel (1991). Their model is described in the next section. Two iterative processes to find the ML estimates of the model parameters are discussed in Section 3. The closed form expressions for the asymptotic covariances of the estimators of model parameters and of the fitted values are provided in Section 4. The published Canadian retail trade data collected by Statistics Canada are used to illustrate the methodology.

2. CONSTANT MULTIPLICATIVE BIAS MODEL (CMBM)

In order to meet the benchmarking requirements of the economic surveys, the following constant multiplicative bias model (CMBM) has been proposed by Laniel and Fyfe (1989; 1990). The model assumes that the biased sub-annual estimates y_t follow the relationship given by

$$y_t = \beta\theta_t + a_t, \quad t = 1, 2, \ldots, n \tag{2.1}$$

and the unbiased annual estimates z_T follow the relationship

$$z_T = \sum_{t \in T} \theta_t + b_T, \quad T = 1, 2, \ldots, m, \tag{2.2}$$

where the subscripts t and T denotes respectively the sub-annual and annual time periods, θ_t is the unknown fixed sub-annual parameter, β is an unknown constant bias parameter associated with y_t, and a_t and b_t are sampling errors associated respectively with y_t and z_T. The above model is a hybrid type (mixed) model in which bias is multiplicative but errors are additive.

Before proceeding further, let us define the column vectors $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$, $\boldsymbol{z} = (z_1, z_2, \ldots, z_m)'$, $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)'$, $\boldsymbol{b} = (b_1, b_2, \ldots, b_m)'$, and $\boldsymbol{\Theta} = (\theta_1, \theta_2, \ldots, \theta_n)'$. The CMBM model, given by (2.1) and (2.2), can be rewritten as

$$\begin{aligned} \boldsymbol{w} &= X_\beta \boldsymbol{\Theta} + \boldsymbol{u} \\ &= X_\Theta \beta + X_D \boldsymbol{\Theta} + \boldsymbol{u}, \end{aligned} \tag{2.3}$$

where

$$X_\beta = (\beta I_n : \boldsymbol{D}')', \quad X_\Theta = (\boldsymbol{\Theta}' : \boldsymbol{0}')', \quad X_D = (\boldsymbol{0}' : \boldsymbol{D}')',$$
$$\boldsymbol{w} = (\boldsymbol{y}' : \boldsymbol{z}')', \quad \boldsymbol{u} = (\boldsymbol{a}' : \boldsymbol{b}')', \quad \boldsymbol{D} = (d_{Tt}), \tag{2.4}$$

I_n is an identity matrix of order n, $\boldsymbol{0}$ is a zero vector or matrix of an appropriate order, and d_{Tt} is an indicator function equal to 1 for $t \in T$ and to 0 otherwise. It is assumed that the sampling error vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ follow multivariate normal distributions such that $\boldsymbol{a} \sim MN(\boldsymbol{0}, V_{aa})$ and $\boldsymbol{b} \sim MN(\boldsymbol{0}, V_{bb})$. Also, in the general case, $\boldsymbol{a}$ and $\boldsymbol{b}$ are correlated, which means that $\text{Cov}(\boldsymbol{a},\boldsymbol{b}) = V_{ab} = V'_{ba} \neq \boldsymbol{0}$. It is shown in the next section that the ML and GLS estimators of the $\boldsymbol{\Theta}$ and β are same for this model. Thus the assumption regarding the normality of $\boldsymbol{a}$ and $\boldsymbol{b}$ is required only to obtain the Fisher information matrix (and hence variances) of the ML estimators.

3. MAXIMUM LIKELIHOOD ESTIMATION

The log-likelihood function under CMBM can be written as

$$\ln(L) = -\frac{(n + m)}{2}\ln(2\pi) - \frac{1}{2}\ln|V| - \frac{1}{2}Q, \tag{3.1}$$

where

$$Q = (\boldsymbol{w} - X_\beta\boldsymbol{\Theta})' V^{-1} (\boldsymbol{w} - X_\beta\boldsymbol{\Theta}) \tag{3.2}$$

and

$$V = \begin{pmatrix} V_{aa} & V_{ab} \\ V_{ba} & V_{bb} \end{pmatrix}.$$

The ML estimates of the model parameters $\boldsymbol{\Theta}$ and β can be obtained, assuming V known, by maximizing the log-likelihood function (3.1) or equivalently by minimizing the quadratic term Q (3.2). For this particular model, the ML and GLS estimators of the model parameters are the same and the distinction between them will be made only if the need arises. Taking the first order partial derivatives of $\ln(L)$ with respect to $\boldsymbol{\Theta}$ and β, respectively, and then equating them to zero, we have

$$\begin{aligned} \frac{\partial \ln(L)}{\partial \boldsymbol{\Theta}} &= X'_\beta V^{-1}(\boldsymbol{w} - X_\beta\boldsymbol{\Theta}) = \boldsymbol{0}, \\ \frac{\partial \ln(L)}{\partial \beta} &= X'_\Theta V^{-1}(\boldsymbol{w} - X_\beta\boldsymbol{\Theta}) = 0. \end{aligned} \tag{3.3}$$

Since $E(w) = X_\beta \mathbf{\Theta}$ under the model (2.3), the above equations are estimating equations in the sense of Godambe (1960) and they are information unbiased. It is interesting to note that $X_\beta' V^{-1}$ and $X_\Theta' V^{-1}$ do not depend on w so that the equations (3.3) converge to zeros and hence have consistent roots as long as $E(w) = X_\beta \mathbf{\Theta}$. That is, even when V in the above equations is replaced by some of its consistent estimate the equations will provide consistent estimates of the vector $\mathbf{\Theta}$ and β. Also note that the above equations are non-linear in the parameters to be estimated and it is not possible to obtain explicit expressions for the estimators of $\mathbf{\Theta}$ and β. Therefore some iterative procedure, such as the well-known Fisher-Newton-Raphson method (also called method of scores by Fisher), may be used to obtain the estimates. The elements of expected Fisher information matrix needed to implement the Fisher-Newton-Raphson method are provided in Section 4.

An alternate way to find the ML estimates of the model parameters is to solve the estimating equations (3.3) successively. By solving the first expression of (3.3), the estimate of $\mathbf{\Theta}$, as a function of β, is given by

$$\hat{\mathbf{\Theta}} \equiv \hat{\mathbf{\Theta}}(\beta) = (X_\beta' V^{-1} X_\beta)^{-1} X_\beta' V^{-1} w. \qquad (3.4)$$

Similarly, by solving the second expression of (3.3), the estimator of β, as a function of $\mathbf{\Theta}$, is given by

$$\hat{\beta} \equiv \hat{\beta}(\mathbf{\Theta}) = \left[\mathbf{\Theta}' V_{aa.b}^{-1}\left(y - V_{ab} V_{bb}^{-1}(z - D\mathbf{\Theta})\right)\right] \Big/ \left[\mathbf{\Theta}' V_{aa.b}^{-1} \mathbf{\Theta}\right], \qquad (3.5)$$

where

$$V_{aa.b} = V_{aa} - V_{ab} V_{bb}^{-1} V_{ba}.$$

The ML estimates of $\mathbf{\Theta}$ and β can be obtained by successively calculating equations (3.4) and (3.5) until convergence. This procedure has an advantage over the Fisher-Newton-Raphson method as it is easy to implement. However, for this kind of algorithm, the convergence is usually very slow. We will compare these two methods in Section 6 to check the speed of their convergence.

Once the ML estimates of the model parameters are obtained, one can find the fitted sub-annual values $\hat{y} = \hat{\beta}\hat{\mathbf{\Theta}}$ and the fitted annual values $\hat{z} = D\hat{\mathbf{\Theta}}$.

Initial Guess for $\mathbf{\Theta}$ and β

In order to obtain an initial guess for β, say $\hat{\beta}_0$, let us rewrite the model (2.3) as

$$w^* = X_\Theta^* \beta + u^*,$$

where $w^* = ((Dy)' : (z - D\mathbf{\Theta})')'$, $X_\Theta^* = ((D\mathbf{\Theta})' : \mathbf{0}')'$ and $u^* = ((Da)' : b')'$. Thus the ML estimate of β is given by

$$\hat{\beta} = \left[X_\Theta^{*\prime} (V^*)^{-1} w^*\right] \Big/ \left[X_\Theta^{*\prime} (V^*)^{-1} X_\Theta^*\right], \qquad (3.6)$$

where

$$V^* = \operatorname{Cov}(u^*) = \begin{pmatrix} DV_{aa}D' & DV_{ab} \\ V_{ba}D' & V_{bb} \end{pmatrix}.$$

Using the fact that $E(z) = D\mathbf{\Theta}$, and replacing $D\mathbf{\Theta}$ by z in (3.6), an initial guess for β may be taken as

$$\hat{\beta}_0 = \left[\begin{pmatrix} z \\ \mathbf{0} \end{pmatrix}' (V^*)^{-1} w^*\right] \Big/ \left[\begin{pmatrix} z \\ \mathbf{0} \end{pmatrix}' (V^*)^{-1} \begin{pmatrix} z \\ \mathbf{0} \end{pmatrix}\right] \qquad (3.7)$$

$$= \left[z'(DV_{aa.b}D')^{-1} Dy\right] \Big/ \left[z'(DV_{aa.b}D')^{-1} z\right].$$

The initial estimate of $\mathbf{\Theta}$ can be obtained from (3.4) by replacing β by $\hat{\beta}_0$.

4. COVARIANCES OF THE ESTIMATORS

In this section, we derive the expressions for the asymptotic covariances of the ML estimators of CMBM parameters by inverting the Fisher information matrix, say $\mathbf{\Omega}$. The asymptotic covariances of the fitted sub-annual and annual values are provided by using the delta method. First, let us consider the derivation of the covariances of the ML estimators of $\mathbf{\Theta}$ and β. The elements of $\mathbf{\Omega}$ (*i.e.*, the negative expectations of the second order partial derivatives of $\ln(L)$) are given by

$$\mathbf{\Omega}_{11} = -E\left[\frac{\partial^2 \ln(L)}{\partial \mathbf{\Theta}\, \partial \mathbf{\Theta}'}\right] = X_\beta' V^{-1} X_\beta,$$

$$\mathbf{\Omega}_{22} = -E\left[\frac{\partial^2 \ln(L)}{\partial \beta^2}\right] = \mathbf{\Theta}' V_{aa.b}^{-1} \mathbf{\Theta}$$

and

$$\mathbf{\Omega}_{12} = \mathbf{\Omega}_{21}' = -E\left[\frac{\partial^2 \ln(L)}{\partial \mathbf{\Theta}\, \partial \beta}\right] = X_\beta' V^{-1} X_\Theta.$$

Therefore, the Fisher information matrix of order $(n + 1) \times (n + 1)$ is given by

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{bmatrix}. \qquad (4.1)$$

Inverting $\boldsymbol{\Omega}$ by using the algebra of partitioned matrices we have

$$\begin{aligned} \operatorname{Cov}(\hat{\boldsymbol{\Theta}}) &= \boldsymbol{\Omega}_{11.2}^{-1}, \\ \operatorname{Var}(\hat{\beta}) &= \boldsymbol{\Omega}_{22.1}^{-1}, \\ \operatorname{Cov}(\hat{\boldsymbol{\Theta}},\hat{\beta}) &= -\,\boldsymbol{\Omega}_{11.2}^{-1}\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1} \\ &= -\,\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22.1}^{-1}, \end{aligned} \tag{4.2}$$

where

$$\begin{aligned} \boldsymbol{\Omega}_{11.2} &= \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}, \\ \boldsymbol{\Omega}_{22.1} &= \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}. \end{aligned} \tag{4.3}$$

Once the covariance matrix $\boldsymbol{\Omega}^{-1}$ is available, the asymptotic covariances of the sub-annual fitted values $\hat{\boldsymbol{y}}$ can be obtained by using the delta method (see *e.g.*, Rao 1973). Let $\boldsymbol{\Delta}$ be the matrix of first order partial derivatives of $\boldsymbol{y}$ with respect to the elements of $(\boldsymbol{\Theta}':\beta)'$. Clearly, the $n \times (n+1)$ matrix is $\boldsymbol{\Delta} = (\beta \boldsymbol{I}_n : \boldsymbol{\Theta})$. Now, by using the delta method, the asymptotic covariance matrix of $\hat{\boldsymbol{y}}$ is given by

$$\operatorname{Cov}(\hat{\boldsymbol{y}}) = \boldsymbol{\Delta}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}'. \tag{4.4}$$

Furthermore, the covariance matrix of the annual fitted values $\hat{\boldsymbol{z}}$, from the standard multivariate normal theory, is given by

$$\operatorname{Cov}(\hat{\boldsymbol{z}}) = \boldsymbol{D}\boldsymbol{\Omega}_{11.2}^{-1}\boldsymbol{D}', \tag{4.5}$$

where $\boldsymbol{D}$ and $\boldsymbol{\Omega}_{11.2}$ are as defined by (2.4) and (4.3), respectively.

5. MAXIMUM LIKELIHOOD ESTIMATION WHEN $\boldsymbol{V}_{ab} = \boldsymbol{0}$

In this section we consider the ML estimation of the model parameters for the special case when the error vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are uncorrelated (*i.e.*, $\operatorname{Cov}(\boldsymbol{a},\boldsymbol{b}) = \boldsymbol{V}_{ab} = \boldsymbol{V}_{ba}' = \boldsymbol{0}$). Usually this is the case in sample surveys when annual and sub-annual samples are drawn independently from each other. Reduction in the results of Sections 3 and 4 can be seen by substituting $\boldsymbol{V}_{ab} = \boldsymbol{V}_{ba}' = \boldsymbol{0}$ in the equations. As an example, for this special case, the ML estimators of $\boldsymbol{\Theta}$ and β, given by (3.4) and (3.5), reduce to

$$\hat{\boldsymbol{\Theta}}^* \equiv \hat{\boldsymbol{\Theta}}^*(\beta) = \left(\beta^2 \boldsymbol{V}_{aa}^{-1} + \boldsymbol{D}'\boldsymbol{V}_{bb}^{-1}\boldsymbol{D}\right)^{-1} \left(\beta \boldsymbol{V}_{aa}^{-1}\boldsymbol{y} + \boldsymbol{D}'\boldsymbol{V}_{bb}^{-1}\boldsymbol{z}\right)$$

and

$$\hat{\beta}^* \equiv \hat{\beta}^*(\boldsymbol{\Theta}) = \left[\boldsymbol{\Theta}'\boldsymbol{V}_{aa}^{-1}\boldsymbol{y}\right] / \left[\boldsymbol{\Theta}'\boldsymbol{V}_{aa}^{-1}\boldsymbol{\Theta}\right],$$

respectively. These equations must be solved successively to obtain the required estimates.

Similarly, the elements of the Fisher information matrix reduce to

$$\begin{aligned} \boldsymbol{\Omega}_{11}^* &= \beta^2 \boldsymbol{V}_{aa}^{-1} + \boldsymbol{D}'\boldsymbol{V}_{bb}^{-1}\boldsymbol{D}, \\ \boldsymbol{\Omega}_{22}^* &= \boldsymbol{\Theta}'\boldsymbol{V}_{aa}^{-1}\boldsymbol{\Theta}, \\ \boldsymbol{\Omega}_{12}^* &= \boldsymbol{\Omega}_{21}^{*\prime} = \beta \boldsymbol{V}_{aa}^{-1}\boldsymbol{\Theta}. \end{aligned}$$

6. AN APPLICATION

Here we present an example using published Canadian retail trade data which results from monthly and annual retail trade surveys conducted by Statistics Canada. The monthly retail trade estimates and their coefficients of variation (CV) are available from the Statistics Canada publication "Retail Trade" (Catalogue No. 63-005 Monthly). There are two types of monthly retail trade estimates, namely preliminary and revised estimates. We use the revised but seasonally unadjusted (raw) estimates for this example. Since the CVs of the revised estimates are not available, the CVs of the preliminary estimates are used to approximate the variances of the revised monthly estimates. The data for the period January 1985 to December 1988 are used in this example. Another difficulty was to find the autocorrelations for monthly retail trade estimates. Based on some monthly retail trade data, Hidiroglou and Giroux (1986) provided the estimates of autocorrelations at lags 1, 3, 6, 9 and 12 for three different kinds of stratum in several provinces of Canada. As an approximation to the autocorrelations of monthly retail trade estimates, the averages of their estimates of autocorrelations for the strata in the Province of Ontario and Standard Industrial Classification Code 60 (Foods, Beverages, and Drug industries) are used. The approximate (averaged) autocorrelations, say $\rho(k)$, are given in Table 1.

Table 1

Approximate Autocorrelations $\rho(k)$ for the Monthly Retail Trade Estimates

Lag k	1	3	6	9	12
$\rho(k)$	0.970	0.940	0.918	0.914	0.962

The method of ordinary least squares and an algorithm of McLeod (1975) for the derivation of theoretical autocorrelations for autoregressive moving-average time series was used to revise the observed autocorrelations. An ARMA $(1,0)(1,0)_{12}$ seasonal multiplicative model was fitted on the five observed autocorrelations by minimizing the sum of squared differences between the observed and theoretical autocorrelations. Then the estimated model parameters and the above mentioned algorithm of McLeod were used to calculate the autocorrelations for all other lags of interest. Given that the ARMA model is correct for theoretical autocorrelations, this approach provides a consistent estimate of the autocorrelation function. These final (revised) approximate autocorrelations for up to 47 lags are given in Table 2 and were used to approximate the covariances for monthly retail trade estimates via multiplication with the standard deviations.

Table 2

Revised Approximate Autocorrelations $\rho^*(k)$ for the Monthly Retail Trade Estimates for up to 47 Lags

Lag k	$\rho^*(k)$	Lag k	$\rho^*(k)$	Lag k	$\rho^*(k)$	Lag k	$\rho^*(k)$
0	1.0000	12	0.9602	24	0.8896	36	0.8100
1	0.9758	13	0.9345	25	0.8647	37	0.7869
2	0.9555	14	0.9126	26	0.8433	38	0.7669
3	0.9391	15	0.8943	27	0.8253	39	0.7501
4	0.9266	16	0.8798	28	0.8107	40	0.7363
5	0.9177	17	0.8687	29	0.7994	41	0.7254
6	0.9126	18	0.8612	30	0.7913	42	0.7176
7	0.9113	19	0.8572	31	0.7864	43	0.7126
8	0.9136	20	0.8567	32	0.7843	44	0.7106
9	0.9196	21	0.8595	33	0.7862	45	0.7114
10	0.9293	22	0.8661	34	0.7909	46	0.7151
11	0.9429	23	0.8760	35	0.7989	47	0.7217

At the time this study was performed, the annual retail trade estimates were only available for years 1985 through 1988. These estimates are available from Statistics Canada publication "Annual Retail Trade" (Catalogue No. 63-223 Annual). The variances of annual retail trade estimates are not available from the literature and have been computed from the actual survey data. The covariances between monthly and annual estimates are zero because the samples of monthly and annual retail trade surveys were drawn independently from each other. The annual retail trade estimates are from dependent samples, thus their covariances are non-zero. But the estimates of covariances are not readily available via regular survey processing and a study would be required to obtain them. Consequently, for the purpose of this example, we assumed that the covariances between annual retail trade estimates are zero.

An interesting question was raised by one of the referees. He asked what will happen when the variances and covariances of survey estimates are not known. This is a difficult problem and cannot be answered so easily. However the model presented assumes these variances and covariances are known. In general, the estimating equations used to find the maximum likelihood estimates need only the consistent estimates of variances and covariances. It is a common practice in benchmarking problems to estimate these variances and covariances from survey data since the theoretical values are never known (see, *e.g.*, Hillmer and Trabelsi 1987).

The computations required for this example are performed by an algorithm written in the GAUSS programming language for micro computers. The initial estimate of β for the iterative process, obtained form (3.7), is given by $\hat{\beta}_0 = 0.9162$. The initial estimate of the parameter vector $\mathbf{\Theta}$ is obtained from (3.4), after replacing β by $\hat{\beta}_0$. Both the Fisher-Newton-Raphson and successive iteration methods, as discussed in Section 3, are used to find the ML estimates of the model parameters. The final ML estimate of β is found to be very close to the initial estimate and is given by $\hat{\beta} = 0.9016$ with CV = 0.0065. It is interesting to note that the Fisher-Newton-Raphson method converged very quickly to a final solution for this example. In fact it converged in only 6 iterations (about 1 minute) for a ten digit precision whereas the successive calculations method converged, with the same precision, in over 500 iterations (over 45 minutes), on a 386DX-25Mhz personal computer. However, as they should, both methods converged to the same final solution. The covariance matrix of the estimated vector $(\hat{\mathbf{\Theta}}':\hat{\beta})'$ is obtained by inverting the Fisher information matrix $\mathbf{\Omega}$, given by (4.1), after replacing parameters by their ML estimates. The original series of the monthly retail trade estimates and the benchmarked series of the ML estimates along with their CVs are given in Table 3. The fitted sub-annual series along with their CVs are also given in this table (last two columns). The original and benchmarked series are also plotted in Figure 1. The results show that the original behaviour of the series is not disturbed by benchmarking and a very large reduction in the CVs of sub-annual estimates is achieved. The original series of the annual retail trade estimates and fitted annual values along with their CVs are given in Table 4. The variances of the fitted values in Tables 3 and 4 are obtained by using expressions (4.4) and (4.5), respectively, after replacing parameters by their ML estimates. The results of fitted values also show a large reduction in the CV's of the original estimates. That is, the reliability of the monthly and annual series are increased by benchmarking.

Table 3

Monthly Retail Trade Estimates, ML Estimates of the Θ_t's and Fitted Values (all in millions of dollars) Along with their CV's

Year	Month	y_t*	CV(y_t)*	$\hat{\Theta}_t$	CV($\hat{\Theta}_t$)	$\hat{y}_t$	CV($\hat{y}_t$)
1985	1	8,689.668	0.008	9,686.630	0.00210	8,733.384	0.00667
	2	8,390.380	0.008	9,350.078	0.00210	8,429.951	0.00665
	3	10,107.485	0.006	11,248.048	0.00233	10,141.146	0.00496
	4	10,541.145	0.008	11,741.785	0.00200	10,586.294	0.00656
	5	11,763.659	0.007	13,094.151	0.00198	11,805.576	0.00570
	6	11,067.487	0.008	12,321.326	0.00189	11,108.803	0.00647
	7	10,810.755	0.008	12,029.467	0.00184	10,845.666	0.00643
	8	11,289.656	0.009	12,554.808	0.00206	11,319.309	0.00726
	9	10,336.540	0.009	11,484.216	0.00205	10,354.073	0.00728
	10	11,213.751	0.010	12,447.696	0.00256	11,222.737	0.00809
	11	11,935.495	0.010	13,234.412	0.00258	11,932.034	0.00808
	12	13,300.288	0.008	14,734.891	0.00188	13,284.853	0.00643
1986	1	9,753.373	0.009	10,794.009	0.00221	9,731.787	0.00716
	2	9,249.279	0.009	10,227.777	0.00224	9,221.277	0.00709
	3	10,609.952	0.008	11,729.293	0.00207	10,575.031	0.00622
	4	11,637.936	0.008	12,860.626	0.00206	11,595.032	0.00614
	5	12,695.108	0.008	14,024.139	0.00205	12,644.046	0.00605
	6	11,826.254	0.008	13,059.556	0.00202	11,774.385	0.00598
	7	11,940.908	0.010	13,164.500	0.00233	11,869.002	0.00740
	8	11,866.547	0.010	13,070.205	0.00232	11,783.987	0.00743
	9	11,540.397	0.009	12,712.283	0.00202	11,461.287	0.00670
	10	12,208.845	0.010	13,430.932	0.00235	12,109.215	0.00747
	11	12,201.498	0.010	13,418.219	0.00240	12,097.753	0.00747
	12	14,479.170	0.009	15,933.951	0.00215	14,365.916	0.00670
1987	1	10,271.723	0.012	11,276.676	0.00357	10,166.956	0.00891
	2	9,951.105	0.010	10,945.319	0.00261	9,868.208	0.00737
	3	11,492.162	0.008	12,663.849	0.00230	11,417.620	0.00584
	4	12,867.443	0.009	14,172.605	0.00235	12,777.901	0.00652
	5	13,508.434	0.012	14,850.145	0.00343	13,388.765	0.00862
	6	13,608.274	0.011	14,973.985	0.00287	13,500.418	0.00786
	7	13,278.474	0.023	14,483.340	0.01066	13,058.057	0.00165
	8	12,728.196	0.008	14,028.998	0.00227	12,648.426	0.00577
	9	12,616.239	0.009	13,888.982	0.00233	12,522.188	0.00659
	10	13,760.829	0.008	15,156.409	0.00227	13,664.890	0.00592
	11	13,380.142	0.008	14,733.240	0.00227	13,283.365	0.00597
	12	16,269.757	0.007	17,928.148	0.00241	16,163.867	0.00525
1988	1	11,134.013	0.010	12,234.529	0.00274	11,030.548	0.00753
	2	10,959.374	0.010	12,042.761	0.00276	10,857.651	0.00754
	3	13,177.788	0.008	14,508.565	0.00233	13,080.800	0.00602
	4	13,666.311	0.009	15,035.737	0.00243	13,556.094	0.00676
	5	14,267.530	0.006	15,742.039	0.00379	14,192.890	0.00448
	6	14,432.944	0.009	15,884.130	0.00240	14,320.997	0.00673
	7	13,960.825	0.009	15,363.957	0.00240	13,852.014	0.00673
	8	13,691.315	0.008	15,073.691	0.00233	13,590.312	0.00606
	9	13,773.109	0.008	15,159.075	0.00235	13,667.294	0.00613
	10	13,900.743	0.009	15,279.950	0.00255	13,776.282	0.00696
	11	14,453.461	0.009	15,884.279	0.00260	14,321.132	0.00700
	12	17,772.990	0.009	19,529.791	0.00267	17,607.895	0.00702

***Source:** Statistics Canada publication "Retail Trade" (Catalogue No. 63-005 Monthly).

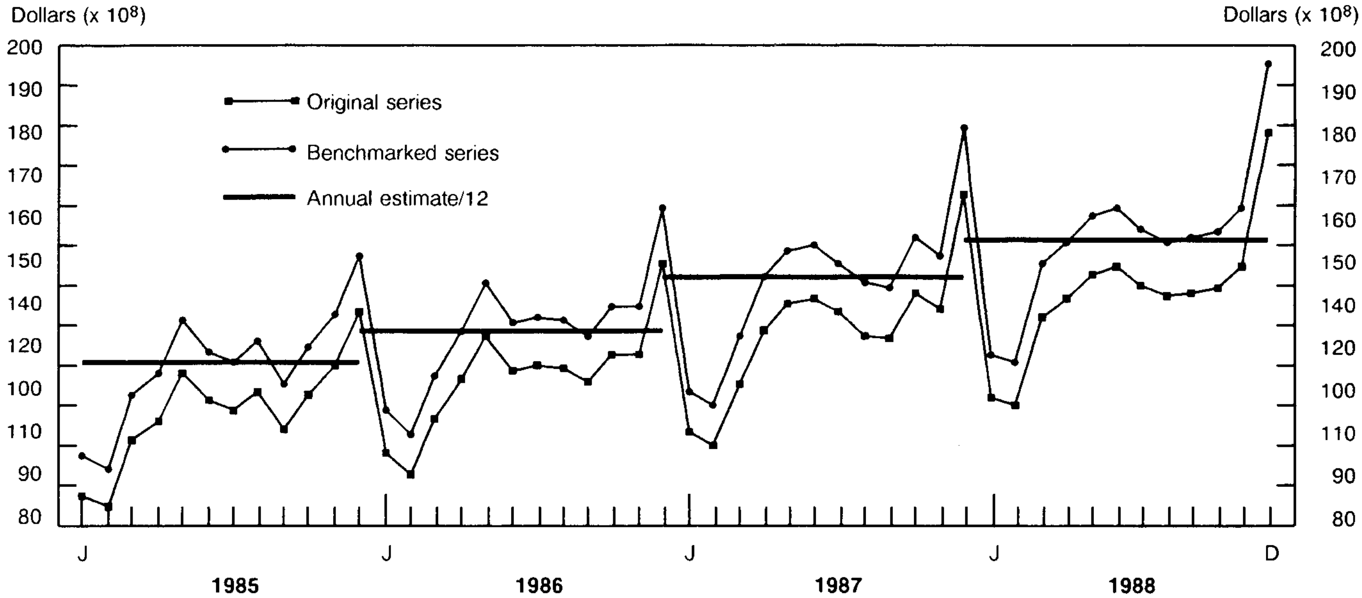

Figure 1. Original and Benchmarked Series of Monthly Retail Trade Estimates for All Stores in Canada

Table 4

Annual Retail Trade Estimates and Annual Fitted Values (in millions of dollars) Along with their CV's

Year	z_T^*	CV(z_T)	$\hat{z}_T$	CV$(\hat{z}_T)$
1985	143,965.400	0.00033	143,927.507	0.00032
1986	154,377.100	0.00031	154,425.491	0.00030
1987	169,944.600	0.00193	169,101.697	0.00128
1988	181,594.000	0.00137	181,738.512	0.00127

* **Source**: Statistics Canada publication "Annual Retail Trade" (Catalogue No. 63-223 Annual).

7. CONCLUSIONS

The non-linear model discussed here seems to be very appropriate for benchmarking an economic time series from large sample surveys. The proposed iterative procedures to find the maximum likelihood estimates of the model parameters are very simple to implement in practice. However, the convergence of the successive calculation method is very slow in comparison to the Fisher-Newton-Raphson method. The closed form expressions for the covariances of the ML estimators are provided. These estimates and their covariances may be used to make inferences regarding model parameters. Furthermore, expressions for the fitted sub-annual and annual values along with their asymptotic covariances are also provided. The methodology presented in this article seems to provide a good fit to the Canadian retail trade data. However, the goodness of fit tests for this and other benchmarking models need to be developed.

ACKNOWLEDGEMENTS

We acknowledge the valuable suggestions made by the editor, associate editor and two referees which had greatly improved the quality of this paper. We are also thankful to Drs. J. Gambino and M. Kovačević of Statistics Canada for their comments on this research.

REFERENCES

BOX, G.E.P., and JENKINS, G.M. (1976). *Time Series Analysis, Forecasting and Control.* New York: Holden-Day.

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

CHOLETTE, P.A. (1987). Benchmarking and interpolation of time series. Methodology Branch Working Paper, Statistics Canada.

CHOLETTE, P.A. (1988). Benchmarking systems of socio-economic time series. Methodology Branch Working Paper, Statistics Canada.

CHOLETTE, P.A. (1992). Users' manual of programmes BENCH and CALEND to benchmark, interpolate and Calendarize time series data on micro computers. Methodology Branch Working Paper, Statistics Canada.

CHOLETTE, P.A., and DAGUM, E.B. (1989). Benchmarking socio-economic time series data: a unified approach. Methodology Branch Working Paper, Statistics Canada.

CHOLETTE, P.A., and DAGUM, E.B. (1991). Benchmarking time series with autocorrelated sampling errors. Methodology Branch Working Paper, Statistics Canada.

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

GODAMBE, V.P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 13, 1208-1211.

HIDIROGLOU, M.A., and GIROUX, S. (1986). Composite estimation for the Retail Trade Survey. Methodology Branch Working Paper, Statistics Canada.

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

LANIEL, N., and FYFE, K. (1989). Benchmarking of economic time series. Methodology Branch Working Paper, Statistics Canada.

LANIEL, N., and FYFE, K. (1990). Benchmarking of economic time series. *Survey Methodology*, 16, 271-277.

LANIEL, N., and MIAN, I.U.H. (1991). Maximum likelihood estimation for the constant bias model with mixed benchmarks. Methodology Branch Working Paper, Statistics Canada.

McLEOD, I. (1975). Derivation of the theoretical autocovariance function of autoregressive-moving average time series. *Applied Statistics*, 24, 255-256.

MIAN, I.U.H., and LANIEL, N. (1991). Maximum likelihood estimation for the constant bias benchmarking model. Methodology Branch Working Paper, Statistics Canada.

RAO, C.R. (1973). *Linear Statistical Inference and Its Applications*, (2nd Ed.). New York: John Wiley.