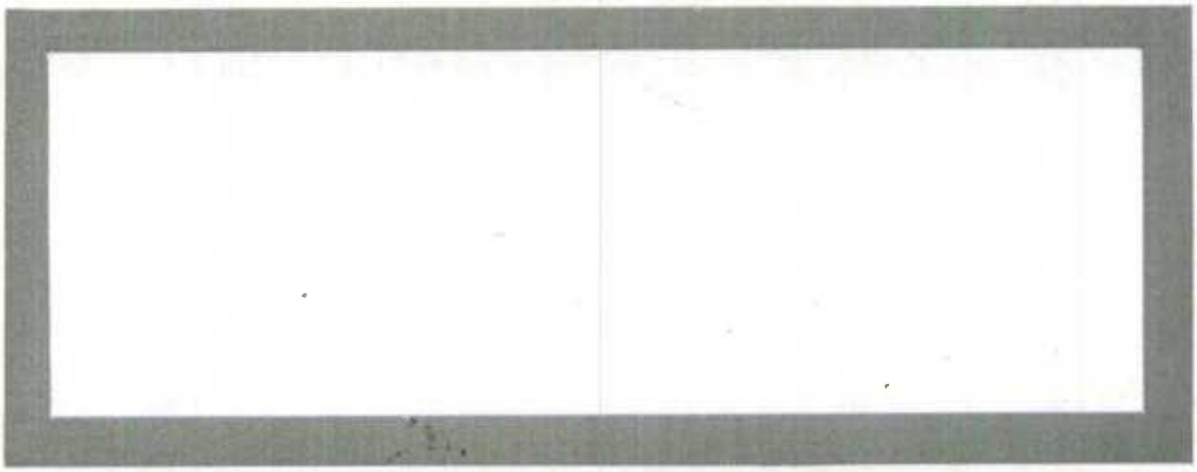
Statistics Canada Statistique Canada



Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes-entreprises

Canada

WORKING PAPER NO. BSMD-98-009E
METHODOLOGY BRANCH

CAHIER DE TRAVAIL NO. BSMD-98-009E
DIRECTION DE LA MÉTHODOLOGIE

Best Feasible Unbiased Prediction for Multi-Source Data System With an Application for Multiplicative Benchmarking

Zhao-Guo Chen and Estela Bee Dagum

October 1998

Best Feasible Unbiased Prediction for Multi-Source Data System with an Application for Multiplicative Benchmarking

by

Zhao-Guo Chen[1] and Estela Bee Dagum[2]

October 1998

Time Series Research and Analysis Centre
Business Survey Methodology Division

Statistics Canada

[1] *Address for correspondence:* Time Series Research and Analysis Centre, Business Survey Methodology Divison, 3-H, R.H.Coats Building, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6. E-mail: chenzha@statcan.ca

[2] *Address for correspondence:* Faculty of Statistical Sciences, University of Bologna, Via delle Belle Arti 41, (40126) Bologna, Italy. E-mail: beedagum@stat.unibo.it

Summary

Information about a socio-economic variable of interest (usually, one or a group of time series) often originates from several sources none of which is complete and/or accurate. A stepwise approach is developed here for predicting the variable of interest by using the data from source to source focusing on minimizing the variances of prediction errors. This paper also reviews some BLUP (the best linear unbiased prediction) theory and shows that the stepwise approach proposed here can give better predictions than BLUP for nonlinear models. As an important application, a nonlinear benchmarking formula for a multiplicative model is derived.

Résumé

L'information sur une variable socio-économique (généralement une ou un ensemble de séries chronologiques) provient souvent de sources différentes, incomplètes ou inexactes. Le présent document présente une approche pas à pas qui prédit la variable en question, en incorporant les donnéees d'une source à la fois, en minimisant la variance des erreurs de prédiction. Cet article passe également en revue la théorie BLUP (meilleure prédiction linéaire non biaisée) et démontre que, pour des modèles non linéaires, l'approche échelonnée proposée peut donner de meilleures prédictions que le BLUP. Un cas d'étalonnage (calage) non linéaire multiplicatif illustre la méthode.

1 Introduction

Suppose that α is the variable of interest and either itself or its functions are observed by different measures with different frequencies and accuracies; that is, several sources of data are available for *predicting* the values of α. Quite often, we consider situations where the values of α form a time series, $\{\alpha(t)\}$, or a group of time series.

We use the term "*prediction*" (rather than "estimation") from the literature of estimating random effects (see e.g. Henderson, 1975; Robinson, 1991) that distinguishes estimating random variables from estimating fixed parameters. However, for convenience, when both fixed and random parameters are involved, we will use "prediction". We will also adopt the acronym BLUP (best linear unbiased prediction; in the sense of minimizing the mean square error of predictions). The BLUP of a random vector, say $\boldsymbol{\alpha}$, from a random vector, say $\mathbf{y}$, is denoted by $\hat{E}(\boldsymbol{\alpha} \mid \mathbf{y})$, i.e. the projection of $\boldsymbol{\alpha}$ on the linear space spanned by $\mathbf{y}$. We reserve the notation $E(\boldsymbol{\alpha} \mid \mathbf{y})$ for the conditional expectation of $\boldsymbol{\alpha}$ given $\mathbf{y}$, i.e. the best unbiased prediction (BUP) of $\boldsymbol{\alpha}$ by $\mathbf{y}$.

Predictions could vary wildly depending on the choice of the sources providing observations for prediction. A good prediction should use data from *all* available sources.

This paper proposes an approach for deriving such predictions which for non-linear models may even be better than the BLUP (using data from all sources). Some practical examples where our approach can be applied are those of benchmarking and reconciliation.

In the case of benchmarking, let $\alpha(t)$ denote the monthly total value of a certain product in month t. From the first source, we have monthly observations $y_1(t)$ of $\alpha(t)$ which are contaminated by survey errors $u_1(t)$ either additively, i.e.

$$y_1(t) = \alpha(t) + u_1(t), \quad t = 1, 2, ..., n_1, \tag{1.1}$$

or multiplicatively, i.e.

$$y_1(t) = \alpha(t)u_1(t), \quad t = 1, 2, ..., n_1. \tag{1.2}$$

In the latter case, $u_1(t)$ represents percentage error; so, the value of $u_1(t)$ should be around 1. Equation (1.2) is an important practical example where the relationship of $\{\alpha(t)\}$ and $\{y_1(t)\}$ is nonlinear.

Assuming that we have some annual indicators or aggregated measures of $\alpha(t)$ from the second source:

$$y_2(k) = \Sigma_{t\in k}\alpha(t) + u_2(k), \quad k = 1, \ldots, n_2, \tag{1.3}$$

where the notation $t \in k$ means the month t is in the year k, and the annual observation errors $u_2(k)$ are much smaller than $u_1(t)$, then $y_2(k), k = 1, \ldots, n_2$, are called benchmarks. A benchmark, $y_2(k)$ is said to be binding if $u_2(k) = 0$ (without error), which is often interpreted as a constraint; otherwise, it is said to be non-binding.

Traditionally, benchmarking consists of using benchmarks to adjust the original measurements or some preliminary predictions of $\boldsymbol{\alpha} = (\alpha(1), \ldots, \alpha(n_1))'$. But it can also consists of using all $y_1(t)$ from the first source and all $y_2(k)$ from the second source to predict $\boldsymbol{\alpha}$ jointly.

Other example is that of reconciliation. Suppose that $\alpha_j(t)$ are the values of a certain industrial product of province j in month t, where $j = 1, ..., J$. At month t, each province has its own observation $y_j(t)$ of $\alpha_j(t)$. However from a different source, we have $y_0(t)$, the observation of the national total of this industry, $\alpha_0(t)$. Naturally, $\alpha_0(t) = \sum_{j=1}^{J} \alpha_j(t)$. But, the corresponding equality does not necessarily hold for $y_j(t), j = 0, 1, ..., J$. The problem is how to give a balanced prediction of the provincial and the national total from these observations so that the equality holds for the predictions for all t. If there are several industries with provincial and national total observations, then balanced predictions are required along two dimensions. If the resulted predictions have also to conform to the annual benchmarks which are from another source, then this problem may be regarded as a three dimensional reconciliation. In the system of national accounts and financial flow areas, statistical agencies often face such problems.

The theory on BLUP is already well developed. To point out some of its limitations in solving the kind of problems mentioned later, we give next a brief review of its

state in the current literature.

Suppose that the linear regression model is as follows,

$$y(t) = \mathbf{x}_f(t)'\boldsymbol{\beta} + \mathbf{x}_r(t)'\boldsymbol{\eta} + e(t), \quad t = 1, \ldots, n_0; \tag{1.4}$$

where $\boldsymbol{\beta}$ is a p unknown vector of fixed numbers (*fixed effects*), while $\boldsymbol{\eta}$ is a q unknown vector of random variables (*random effects*) with $E\boldsymbol{\eta} = 0$; $\mathbf{x}_f(t) = (x_{f1}(t) \ldots x_{fp}(t))'$ and $\mathbf{x}_r(t) = (x_{r1}(t) \ldots x_{rq}(t))'$ are constant vectors; the subscripts f and r stand for "fixed" and "random" respectively. If both $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ exist, the model is called *mixed*. We may write the combination of (1.1) and (1.3) as (1.4) if $\alpha(t) = \mathbf{x}_f(t)'\boldsymbol{\beta} + \eta(t)$, i.e., if the mean of $\alpha(t)$ can be written in linear form $\mathbf{x}_f(t)'\boldsymbol{\beta}$. The prediction of $\alpha(t)$ is then obtained directly from the predictions of $\boldsymbol{\beta}$ and $\eta(t)$.

Using matrix notation, (1.4) can be written as

$$\mathbf{y} = X_f\boldsymbol{\beta} + X_r\boldsymbol{\eta} + \mathbf{e}. \tag{1.5}$$

In the following, the variance matrix of the random vector $\boldsymbol{\eta}$ is denoted by V_η; similarly, for the notation V_e and so on.

For model (1.5), if $p + q < n_0$, then a linear prediction of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ is obtained by using either "joint likelihood" (e.g. Henderson, 1950) or a Bayesian approach (e.g. Dempfle, 1977) under the assumption of normality for the joint distribution of $\boldsymbol{\eta}$ and $\mathbf{e}$ (or prior normality for $\boldsymbol{\eta}$). The solution is obtained from the following equations:

$$\begin{aligned} X_f'V_e^{-1}X_f\hat{\boldsymbol{\beta}} + X_f'V_e^{-1}X_r\hat{\boldsymbol{\eta}} &= X_f'V_e^{-1}\mathbf{y}, \\ X_r'V_e^{-1}X_f\hat{\boldsymbol{\beta}} + (X_r'V_e^{-1}X_r + V_\eta^{-1})\hat{\boldsymbol{\eta}} &= X_r'V_e^{-1}\mathbf{y}. \end{aligned} \tag{1.6}$$

The variance matrix of the prediction errors can also be obtained (see, e.g. Robinson, 1990, p.16). Other authors (e.g. Goldberger, 1962; Harville, 1990) showed directly that this prediction is the BLUP of the parameters without assuming normality. If the assumption of normality holds, then it is the *best unbiased prediction* (BUP). A good review with application examples are given in Robinson (1991).

When $\boldsymbol{\eta}$ is absent, the solution of (1.6) and the variance matrix of the prediction error coincides with the well-known results of generalized least squares (GLS) estimation. When $\boldsymbol{\beta}$ is absent, or $\boldsymbol{\beta}$ is known, the solution of (1.6) and error variance

matrix coincides with the result of the well-known formulae of projection (see, e.g., Whittle, 1963, Section 4.3):

$$\hat{\eta} = \hat{E}(\eta \mid \mathbf{y} - X_f\beta) = V_\eta X_r'(X_r V_\eta X_r' + V_e)^{-1}(\mathbf{y} - X_f\beta)$$

$$= (X_r' V_e^{-1} X_r + V_\eta^{-1})^{-1} X_r' V_e^{-1}(\mathbf{y} - X_f\beta); \tag{1.7}$$

$$E(\hat{\eta} - \eta)(\hat{\eta} - \eta)' = V_\eta - V_\eta X_r'(X_r V_\eta X_r' + V_e)^{-1} X_r V_\eta = (X_r' V_e^{-1} X_r + V_\eta^{-1})^{-1}. \tag{1.8}$$

Here, notice that $\mathbf{y} - X_f\beta = X_r\eta + \mathbf{e}$. In general, β is unknown; the solution for (1.6) is (1.7) but with β replaced by (see, e.g. Robinson, 1990, p.21)

$$\hat{\beta} = [X_f'(X_r V_\eta X_r' + V_e)^{-1} X_f]^{-1} X_f'(X_r V_\eta X_r' + V_e)^{-1}\mathbf{y}. \tag{1.9}$$

The development of the theory and applications of the method of estimating function has a great impact to the problems of estimating and predicting parameters. In this theory, the criterion for deriving targeted estimates, proposed by Godambe (1960), is to minimize a specially defined non-negative matrix. This method of estimating fixed effects (Godambe and Thompson, 1989) can be easily generalized to predict the mixed linear model (1.4) giving the same results as those from (1.6) (see Singh, 1995). The method of estimating function may also be used to more general linear models and even nonlinear models. Singh (1995) provided a good review on this topic and pointed out that, for linear models with bounded number of parameters (in the above case, it means $p+q$ bounded when $n_0 \to \infty$), Godambe's criterion can be interpreted as the minimization of the mean square error of the estimation (prediction) within the class of *linear* unbiased estimating (predicting) function. For nonlinear models with bounded number of parameters, it is the minimization of the *asymptotic* mean square error within the class of *linear* unbiased estimating (predicting) function. Hence, estimating (predicting) function leads to BLUP or asymptotic BLUP for either linear or nonlinear models.

Our approach is not restricted to linear predictions: the minimization of the mean square error of the prediction is within a much larger context. So, it may give better predictions than BLUP. This "better" holds strictly for nonlinear models, such as the

combination of (1.2) and (1.3). Moreover, we do not require the number of parameters in the model to be bounded, it can be as large as the number of observations.

The outline of this paper is as follows. Section 2 states the basic theorem, explains its features and the general procedure for application. Section 3 addresses the multiplicative benchmarking problem where nonlinear benchmarking formulae are derived (providing better predictions than the BLUP). Finally the proof of the basic theorem is presented in the Appendix.

2 The Basic Theorem and the General Procedure

Denote the values of the variable of interest by $\boldsymbol{\alpha} = (\alpha(1), \ldots, \alpha(n))'$ (random vector). Suppose a data system is made up of two sources which offer vectors of observations $\mathbf{y}_1 = (y_1(1), \ldots, y_1(n_1))'$ and $\mathbf{y}_2 = (y_2(1), \ldots, y_2(n_2))'$ respectively.

Sometimes, for the first source, a linear model like (1.1) is good to describe the relationship between $\boldsymbol{\alpha}$ and $\mathbf{y}_1$, which in general can be written as

$$\mathbf{y}_1 = X_1\boldsymbol{\alpha} + \mathbf{u}_1; \tag{2.1}$$

where $\mathbf{u}_1$, the error vector, with mean zero is uncorrelated with $\boldsymbol{\alpha}$. Notice that, unlike (1.1), n_1 is not necessarily the same as n. In general, the relationship between $\boldsymbol{\alpha}$ and $\mathbf{y}_1$ is allowed to be nonlinear, such as in (1.2), which is the case when data have a stable coefficient of variation (CV) rather than a stable standard deviation (SD).

For the second source, we always assume that $\boldsymbol{\alpha}$ and $\mathbf{y}_2$ follow

$$\mathbf{y}_2 = X_2\boldsymbol{\alpha} + \mathbf{u}_2, \tag{2.2}$$

where $\mathbf{u}_2$ has mean zero and is uncorrelated with $\boldsymbol{\alpha}$, and also with $\mathbf{u}_1$ if (2.1) holds (the assumption for more general situations is stated in Theorem 2.1). Here we do not have to specify the distributions of $\mathbf{u}_1$ and $\mathbf{u}_2$, and we allow the variance matrix of $\mathbf{u}_2$ to be degenerate, even 0 (the case where all benchmarks are binding). For convenience, by extending the terminology in benchmarking, we call the kth row in (2.2) a *binding constraint* if $u_2(k) = 0$, otherwise, *non-binding constraint.*

In the following, the dimension of $\boldsymbol{\alpha}$ is always denoted by n; if there are more sources, the dimension of $\mathbf{y}_i$ (the observations from ith source) is denoted by n_i. The notation ":=" means "defined by" or "denoted by".

Definition 2.1 *We call $\hat{\boldsymbol{\alpha}}$ a feasible unbiased prediction (FUP) of $\boldsymbol{\alpha}$ from a source which offers $\mathbf{y}$, the vector of observations (or we call it a FUP of $\boldsymbol{\alpha}$ from $\mathbf{y}$), if $\hat{\boldsymbol{\alpha}}$ is a measurable function of $\mathbf{y}$, $E\hat{\mathbf{e}} = 0$ ($\hat{\mathbf{e}} := \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}$), and $V_{\hat{e}}$ can be obtained.*

A BLUP or a BUP is not necessarily feasible because, sometimes, $V_{\hat{e}}$ can not be obtained either by analytical expressions or by numerical approaches (say, a recursive procedure).

Theorem 2.1 (1) *Let $\hat{\boldsymbol{\alpha}}^{(1)}$ and $\tilde{\boldsymbol{\alpha}}^{(1)}$ be two FUPs of $\boldsymbol{\alpha}$ from the first source, and assume $\mathbf{u}_2$ has mean zero and is uncorrelated with $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}^{(1)}$ and $\tilde{\boldsymbol{\alpha}}^{(1)}$. Denote the prediction-error vectors by $\hat{\mathbf{e}}^{(1)} = \hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha}$ and $\tilde{\mathbf{e}}^{(1)} = \tilde{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha}$. Suppose that we have the inequality*

$$V_{\hat{e}^{(1)}} \leq V_{\tilde{e}^{(1)}} \tag{2.3}$$

and that

$$X_2 V_{\hat{e}^{(1)}} X_2' + V_{u_2} > 0, \tag{2.4}$$

holds. Let

$$\hat{\boldsymbol{\alpha}}^{(2)} := \hat{\boldsymbol{\alpha}}^{(1)} + V_{\hat{e}^{(1)}} X_2' (X_2 V_{\hat{e}^{(1)}} X_2' + V_{u_2})^{-1} (\mathbf{y}_2 - X_2 \hat{\boldsymbol{\alpha}}^{(1)}), \tag{2.5}$$

and a similar formula for $\tilde{\boldsymbol{\alpha}}^{(2)}$ from $\tilde{\boldsymbol{\alpha}}^{(1)}$ and $V_{\tilde{e}^{(1)}}$. Let $\hat{\mathbf{e}}^{(2)} = \hat{\boldsymbol{\alpha}}^{(2)} - \boldsymbol{\alpha}$ and $\tilde{\mathbf{e}}^{(2)} = \tilde{\boldsymbol{\alpha}}^{(2)} - \boldsymbol{\alpha}$, then

$$V_{\hat{e}^{(2)}} = V_{\hat{e}^{(1)}} - V_{\hat{e}^{(1)}} X_2' (X_2 V_{\hat{e}^{(1)}} X_2' + V_{u_2})^{-1} X_2 V_{\hat{e}^{(1)}}, \tag{2.6}$$

similar relationship for $V_{\tilde{e}^{(2)}}$ and $V_{\tilde{e}^{(1)}}$ holds, and

$$V_{\hat{e}^{(2)}} \leq V_{\tilde{e}^{(2)}}. \tag{2.7}$$

(2) *Assume a third source which offers observations $\mathbf{y}_3 = X_3 \boldsymbol{\alpha} + \mathbf{u}_3$ is available, where $\mathbf{u}_3$ has mean zero and is uncorrelated with $\mathbf{u}_2, \boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}^{(1)}$. Let*

$$\bar{\mathbf{y}}_3 = \begin{pmatrix} \mathbf{y}_2 \\ \mathbf{y}_3 \end{pmatrix}, \quad \bar{X}_3 = \begin{pmatrix} X_2 \\ X_3 \end{pmatrix}, \quad \bar{\mathbf{u}}_3 = \begin{pmatrix} \mathbf{u}_2 \\ \mathbf{u}_3 \end{pmatrix}, \tag{2.8}$$

then

$$\bar{X}_3 V_{\hat{e}^{(1)}} \bar{X}_3' + V_{\bar{u}_3} > 0. \tag{2.9}$$

holds if and only if both (2.4) and

$$X_3 V_{\hat{e}^{(2)}} X_3' + V_{u_3} > 0 \tag{2.10}$$

hold. Under condition (2.9),

$$\hat{\boldsymbol{\alpha}}^{(3)} := \hat{\boldsymbol{\alpha}}^{(1)} + V_{\hat{e}^{(1)}} \bar{X}_3' (\bar{X}_3 V_{\hat{e}^{(1)}} \bar{X}_3' + V_{\bar{u}_3})^{-1} (\bar{\mathbf{y}}_3 - \bar{X}_3 \hat{\boldsymbol{\alpha}}^{(1)}) \tag{2.11}$$

$$= \hat{\boldsymbol{\alpha}}^{(2)} + V_{\hat{e}^{(2)}} X_3' (X_3 V_{\hat{e}^{(2)}} X_3' + V_{u_3})^{-1} (\mathbf{y}_3 - X_3 \hat{\boldsymbol{\alpha}}^{(2)}), \tag{2.12}$$

and the variance matrix of its prediction-error vector $\hat{\mathbf{e}}^{(3)} = \hat{\boldsymbol{\alpha}}^{(3)} - \boldsymbol{\alpha}$ *is*

$$V_{\hat{e}^{(3)}} = V_{\hat{e}^{(1)}} - V_{\hat{e}^{(1)}} \bar{X}_3' (\bar{X}_3 V_{\hat{e}^{(1)}} \bar{X}_3' + V_{\bar{u}_3})^{-1} \bar{X}_3 V_{\hat{e}^{(1)}}$$

$$= V_{\hat{e}^{(2)}} - V_{\hat{e}^{(2)}} X_3' (X_3 V_{\hat{e}^{(2)}} X_3' + V_{u_3})^{-1} X_3 V_{\hat{e}^{(2)}}. \tag{2.13}$$

(3) *If* $\hat{\boldsymbol{\alpha}}^{(1)} = \hat{E}(\boldsymbol{\alpha} \mid \mathbf{y}_1)$*, the BLUP of* $\boldsymbol{\alpha}$ *from* $\mathbf{y}_1$*, is also a FUP of* $\boldsymbol{\alpha}$ *and (2.4) holds, then* $\hat{\boldsymbol{\alpha}}^{(2)}$ *given by (2.5) is* $\hat{E}(\boldsymbol{\alpha} \mid \mathbf{y}_1, \mathbf{y}_2)$*, the BLUP of* $\boldsymbol{\alpha}$ *jointly from* $\mathbf{y}_1$ *and* $\mathbf{y}_2$*. As* $V_{\hat{e}^{(2)}}$ *is given by (2.6), it is also a FUP.*

(4) *Suppose that* $\hat{\boldsymbol{\alpha}}^{(1)} = E(\boldsymbol{\alpha} \mid \mathbf{y}_1)$*, the BUP of* $\boldsymbol{\alpha}$ *from* $\mathbf{y}_1$*, is also feasible with variance matrix of prediction errors* $V_{\hat{e}^{(1)}}$*, and that* $\mathbf{u}_2$ *has mean zero and is independent of* $\boldsymbol{\alpha}$ *and* $\mathbf{y}_1$*. If (2.4) holds and* $\hat{\boldsymbol{\alpha}}^{(2)}$ *is given by (2.5). Define the set*

$$\mathcal{A} = \{\tilde{\boldsymbol{\alpha}} : \ \tilde{\boldsymbol{\alpha}} = \mathbf{a}(\mathbf{y}_1) + A_2 \mathbf{y}_2\}, \tag{2.14}$$

where $\mathbf{a}(\mathbf{y}_1)$ *is a vector of dimension* n *of measurable functions of* $\mathbf{y}_1$*, and* A_2 *is an* $n \times n_2$ *matrix of constants. Then* $V_{\hat{e}^{(2)}}$ *given by (2.6), the variance matrix of* $\hat{\mathbf{e}}^{(2)} = \hat{\boldsymbol{\alpha}}^{(2)} - \boldsymbol{\alpha}$*, minimizes* $V_{\tilde{e}}$ *among all* $\tilde{\mathbf{e}} = \tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}$*,* $\tilde{\boldsymbol{\alpha}} \in \mathcal{A}$*.*

Remark 2.1 Suppose that $\hat{\boldsymbol{\alpha}}^{(1)}$ is a FUP from the first source and there are more sources available which give $\mathbf{y}_i = X_i \boldsymbol{\alpha} + \mathbf{u}_i$, $i = 2, \ldots, m$. All $\mathbf{u}_i$ are mutually uncorrelated and uncorrelated with $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}^{(1)}$ (equivalently, $\hat{\mathbf{e}}^{(1)}$). Let $\bar{\mathbf{y}}_m$, $\bar{X}_m$ and $\bar{\mathbf{u}}_m$ be defined similar to (2.8), and a condition like (2.9) holds; then we may obtain $\hat{\boldsymbol{\alpha}}^{(m)}$ directly from (2.5) with X_2 and V_{u_2} replaced by $\bar{X}_m$ and $V_{\bar{u}_m}$. However, the

second part of the theorem tells that, rather than pooling all these sources together as one, a more preferable way is using (2.5) with (2.6) as a recursive formulae (replacing 1 by $i-1$ and 2 by i, $i=2,\ldots,m$), inputting data from source to source. We finally get the same result which is independent of the order followed by the sources entering in the recursive procedure. The dimensions of the matrices for inversion in the recursive procedure may become much smaller.

We can even split some originally defined sources into a series of newly defined "sources" provided that the new $\mathbf{u}_i$ are uncorrelated with each other (binding constraints, $\mathbf{u}_i = 0$, automatically fulfill this condition). This is particularly useful when there is a huge number of constraints (e.g. national account systems).

Another advantage is that we can monitor whether a new single constraint or a group of constraints is redundant. Chen and Dagum (1998)show that there are no redundant constraints until step j if and only if $X_i V_{\hat{e}^{(i)}} X_i' + V_{u_i} > 0$ holds for all $i \leq j$. Notice that redundant constraints may turn to be conflicting such that there is no solution for the problem under investigation if no measure of rectification is taken.

Remark 2.2 For classical linear regression models (with fixed effects only), several recursive procedures have been developed under different assumptions. These procedures can be classified as follows: adding a new observation (maybe, a vector) in each step (e.g. Odell and Lewis, 1971; Chambers, 1971), adding a parameter in each step (for the purpose of selecting regressors; see, e.g. Chen and Ni, 1989), and the combination of both (e.g. Jones, 1970).

If the first source is also linear and the global model (using information from all sources) can be written as (1.5) [$\boldsymbol{\alpha}$ is re-parameterized to $(\boldsymbol{\beta}', \boldsymbol{\eta}')'$], then to obtain the BLUP of $\boldsymbol{\alpha}$ using (2.5) with (2.6) (starting with a BLUP from a source) and using (1.7) with (1.9) for predicting $(\boldsymbol{\beta}', \boldsymbol{\eta}')'$ should lead to equivalent results. The first procedure corresponds to adding new observations; the second procedure corresponds to adding new parameters. All of these can be regarded as some extensions of the recursive approach for classical linear regression models (now, for the random effect models or mixed models).

It seems that there are no special advantages by using (1.7) and (1.9) for multiple-

source data systems.

On the other hand, the format of (2.5) and (2.6) is well-known (the format is also similar to the Kalman filtering formulae), and as we mentioned above, the idea of recursion based on entering new observations is fundamental in linear regression and the Kalman filtering theory. However, it had been ignored in the literature that (2.5) with (2.6) can also be used as recursive formulae for entering data, group by group. Such use is more meaningful and powerful when dealing with data systems made from multiple sources.

Noticing the special feature of such data systems, Singh and Kovacevic (1996) proposed segmented Kalman filtering to additive multiple-source benchmarking models. These authors' procedure should lead to an equivalent prediction obtained by recursively using (2.5) and (2.6). However, we believe that the avenue of applying (2.5) and (2.6) as recursive formulae is more direct and has more flexibility for modelling data in practice, especially, when the linear model is not appropriate for the data from the first source.

Remark 2.3 In the literature, some authors have derived formulae such as (2.5) and (2.6) based on different criteria. A typical derivation is based on conditional expectation under the assumption of normality and also of linearity for the first source (e.g. Hillmer and Trabelsi, 1989; Durbin and Quenneville, 1997). In this case, the BLUP and the BUP are the same. Other authors derived the formulae (many of them only discuss the special case $V_{u_2} = 0$) based on constrained maximization of likelihood function (e.g. Weale, 1992). Cholette and Dagum (1994) got their formulae aiming at minimizing the variances of estimation errors under the assumption that $\boldsymbol{\alpha}$ is a vector of constants (fixed effects).

In Theorem 2.1, we directly adopt (2.5) and (2.6) rather than derive them on the basis of a given criterion. According to this theorem, our theoretical goal is to derive the BUP of $\boldsymbol{\alpha}$ from $\mathbf{y}_1$, where their elements can be nonlinear functions of $\mathbf{y}_1$. If this BUP is also feasible, then we use (2.5) and (2.6) to revise this predictions from further sources, and finally obtain the BFUP of $\boldsymbol{\alpha}$ which is defined as follows.

Definition 2.2 *If* $\hat{\boldsymbol{\alpha}}^{(1)} = E(\boldsymbol{\alpha} \mid \mathbf{y}_1)$ *is feasible and the property of independence about* $\bar{\mathbf{u}}_m$, *as* $\mathbf{u}_2$ *in Theorem 2.1 (4), holds; also, condition (2.9) (with 3 replaced by m) holds; then* $\hat{\boldsymbol{\alpha}}^{(m)}$ *initiated from this* $\hat{\boldsymbol{\alpha}}^{(1)}$, *obtained either in one or in several steps using (2.5) and (2.6), is called the best feasible unbiased prediction (BFUP) from all the sources.*

As an important example, we will demonstrate, in the next section, how to obtain a BFUP. The BFUP is the best among the class of predictions defined by (2.14) [with $\mathbf{y}_2$ replaced by $\bar{\mathbf{y}}_m$ and A_2 is an $n \times (n_2 + \ldots + n_m)$ matrix], and is strictly better than the BLUP from these m sources when $\boldsymbol{\alpha}$ is nonlinearly related to $\mathbf{y}_1$).

On the other hand, in the case that we can not obtain the BLUP, then the optimal choice from the first source (under the given circumstance) together with the recursive procedure lead to the best prediction from all the sources for this situation.

Remark 2.4 A key step in the procedure suggested by this paper is to obtain a preliminary prediction, $\hat{\boldsymbol{\alpha}}^{(1)}$, from the first source with as small $V_{\hat{e}^{(1)}}$ as possible. When the relationship of $\boldsymbol{\alpha}$ and $\mathbf{y}_1$ is linear, usually we are satisfied with $\hat{\boldsymbol{\alpha}}^{(1)}$ being the BLUP from $\mathbf{y}_1$. When the number of observations (the dimension of $\mathbf{y}_1$) offered by the first source is at most the same as the number of "parameters" (the dimension of $\boldsymbol{\alpha}$ or the dimension of its re-parmeterization), some special procedures are developed to obtain the BLUP of $\boldsymbol{\alpha}$ from $\mathbf{y}_1$. These procedures are called *signal extraction.*

Put (1.1) in the vector form, i.e. $\mathbf{y}_1 = \boldsymbol{\alpha} + \mathbf{u}_1$. Under the assumption of normality, Hillmer and Trabelsi (1987) derived $\hat{\boldsymbol{\alpha}}^{(1)} = E(\boldsymbol{\alpha} \mid \mathbf{y}_1)$ as follows

$$\begin{cases} \hat{\boldsymbol{\alpha}}^{(1)} &= V_{\hat{e}^{(1)}}(V_{u_1}^{-1}\mathbf{y}_1 + V_{\alpha}^{-1}E\boldsymbol{\alpha}), \\ V_{\hat{e}^{(1)}} &= (V_{u_1}^{-1} + V_{\alpha}^{-1})^{-1}. \end{cases} \tag{2.15}$$

It can be shown that (2.15) is $\hat{E}(\boldsymbol{\alpha} \mid \mathbf{y}_1)$ provided that the first and the second moments exist (see Chen, Cholette and Dagum, 1996); i.e. when normality holds, (2.15) yields $E(\boldsymbol{\alpha} \mid \mathbf{y}_1) = \hat{E}(\boldsymbol{\alpha} \mid \mathbf{y}_1)$ and the covariance matrix of the prediction error.

For implementing a signal extraction formula such as (2.15), often, a model for $\alpha(t)$ is assumed. In other words, $\alpha(t)$ is "re-parameterized". (The number of "parameters"

is even larger than the number of observations.)

A very general scheme of model setting is

$$\alpha(t) = \mu(t) + \eta(t). \tag{2.16}$$

Where the "stochastic component" $\eta(t)$ follows

$$\nabla^d \nabla_s^{\tilde{d}} \eta(t) = -\zeta(t), \tag{2.17}$$

where ∇ and ∇_s are ordinary and seasonal differencing operators respectively and $\{\zeta(t)\}$ is a stationary series with mean zero. The "deterministic part" is expressed by

$$\mu(t) = \sum_{I=1}^{k} z_j(t)\beta_j = \mathbf{z}(t)'\boldsymbol{\beta} \tag{2.18}$$

with known vector-function $\mathbf{z}(t) = (z_1(t) \dots z_k(t))'$ and unknown parameters $\boldsymbol{\beta} = (\beta_1 \dots \beta_k)'$. For some practical considerations of (2.18) in the case of benchmarking and repeated survey, see Chen, Cholette and Dagum (1997).

For $\mu(t) = 0$, (2.16) reduces to a difference stationary (DS) model, and for $\eta(t) = \zeta(t)$, i.e. $d = \tilde{d} = 0$, to a trend stationary (TS) model (Nelson and Plosser, 1982). In model (2.16) through (2.18), the "parameters" are $(\beta_1, \dots, \beta_k, \eta(1), \dots, \eta(n))$ to be predicted from $\mathbf{y}_1$.

We may also cast $\alpha(t)$ in a state space model (see, e.g., Durbin and Quenneville, 1997), and then we have more "parameters" to predict. Kalman filtering and smoothing then can be applied to do the signal extraction.

In model (1.1) if we assume $\alpha(t)$ to be given by (2.16) through (2.18), where denote $d^* = d + s\tilde{d}$, $\boldsymbol{\eta} = (\eta(1) \dots \eta(n))'$, $\boldsymbol{\zeta} = (\zeta(d^*+1) \dots \zeta(n))'$, $Z = (\mathbf{z}(1) \dots \mathbf{z}(n))'$, and D being a $(n - d^*) \times n$ matrix which transfers $\boldsymbol{\eta}$ to $-\boldsymbol{\zeta}$ (see Chen *et al*, 1997), then model (1.1) can be expressed by $\mathbf{y}_1 = \boldsymbol{\alpha} + \mathbf{u}_1$ and put into matrix form as follows:

$$\begin{pmatrix} \mathbf{y}_1 \\ 0 \end{pmatrix} = \begin{pmatrix} Z & I_n \\ 0 & D \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\eta} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_1 \\ \boldsymbol{\zeta} \end{pmatrix}. \tag{2.19}$$

The covariance matrix $V_{u_1} > 0$ and $V_\zeta > 0$ are assumed to be known and Cov $(\mathbf{u}_1, \boldsymbol{\zeta}) = 0$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}' \ \boldsymbol{\eta}')'$, etc, we may simply express (2.19) by

$$\mathbf{y} = X\boldsymbol{\theta} + \mathbf{u}; \tag{2.20}$$

where X is assumed to be of full rank.

Notwithstanding that (2.20) is not a standard regression model (because we impose zeroes on the second part of $\mathbf{y}$, and $\boldsymbol{\zeta}$ is correlated with $\boldsymbol{\eta}$), Dagum, Cholette and Chen (1997) show that, the GLS formulae,

$$\hat{\boldsymbol{\theta}} = V_{\hat{\theta}-\theta} X' V_u^{-1} \mathbf{y}, \quad V_{\hat{\theta}-\theta} = (X' V_u^{-1} X)^{-1}, \tag{2.21}$$

give the BLUP of $\boldsymbol{\theta}$ and the covariance matrix of prediction error respectively; and those of $\boldsymbol{\alpha}$ follow directly from (2.16) and (2.18). Then, the recursive formulae (2.5) and (2.6) can be used for revising the prediction from further sources.

Given the covariance matrix of the survey error, V_{u_1}, by survey experts, a method of estimating V_ζ has been developed by Chen *et al* (1997), so, an estimate of V_u can be obtained from the data.

Remark 2.5 In some cases, due to the nature of the first source, some fixed effects are not predictable by $\mathbf{y}_1$ alone. For example, if the survey error $u_1(t)$ in (1.1) has a bias b, then we can not distinguish b from the level of $\alpha(t)$. A way to overcome this difficulty is as follows. First, we reduce the original "parameters" to a smaller number of new "parameters". For the above example, we regard b as a part of the level of $\alpha(t)$ (e.g., see Durbin and Quenneville, 1997). We can thus get a prediction of these new "parameters" from $\mathbf{y}_1$ with variance matrix of prediction error as small as possible. Second, we obtain a preliminary prediction of the original parameters by combining this prediction of new "parameters" and some observations from further sources with the original "parameters" [see (2.22)] in a linear regression model and carrying out the estimation procedure.

Symbolically, let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1' \ \boldsymbol{\alpha}_2')'$ and suppose that the linear combination $\boldsymbol{\alpha}^* = X_{11}\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ has a FUP from $\mathbf{y}_1$ (for example, $\boldsymbol{\alpha}_1 = b$ and X_{11} is a column of elements of 1); i.e., from the first source we may obtain an unbiased prediction $\hat{\boldsymbol{\alpha}}^* = X_{11}\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + \hat{\mathbf{e}}^*$ of $\boldsymbol{\alpha}^*$ and $V_{\hat{e}^*}$. Pooling this prediction with $\mathbf{y}_2$ (it can be part of observations from a further source), we have

$$\begin{pmatrix} \hat{\boldsymbol{\alpha}}^* \\ \mathrm{y}_2 \end{pmatrix} = \begin{pmatrix} X_{11} & I_{n_1} \\ 0 & X_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix} + \begin{pmatrix} \hat{\mathbf{e}}^* \\ \mathbf{u}_2 \end{pmatrix}. \tag{2.22}$$

If $\mathbf{u}_2$ is uncorrelated with $\hat{\mathbf{e}}^*$ and $\boldsymbol{\alpha}$, and the condition $X_{22}V_{\hat{e}^*}X_{22}'+V_{u_2} > 0$ holds, then based on (2.22), we may apply GLS to obtain the prediction of $\boldsymbol{\alpha}$ and the covariance matrix of the prediction. For the details on the formulae, see Dagum *et al* (1998).

After getting a preliminary prediction of $\boldsymbol{\alpha}$ (denoted by $\hat{\boldsymbol{\alpha}}^{(2)}$ since a second source is involved) and the covariance matrix of the prediction error, the recursive formulae (2.5) and (2.6) can be used for further improvement on the prediction from unused sources.

For two preliminary predictions of $\boldsymbol{\alpha}^*$, $\hat{\boldsymbol{\alpha}}^*$ and $\tilde{\boldsymbol{\alpha}}^*$, if $V_{\hat{e}^*} \leq V_{\tilde{e}^*}$ holds, then (2.7) holds for the corresponding two preliminary predictions of $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}^{(2)}$ and $\tilde{\boldsymbol{\alpha}}^{(2)}$. The proof is similar to that of part (1) in the Appendix and, therefore, is omitted. However, notice that, if $\hat{\boldsymbol{\alpha}}^*$ is the BLUP of $\boldsymbol{\alpha}^*$ from the first source, the approach in the Appendix does not enable one to prove that $\hat{\boldsymbol{\alpha}}^{(2)}$ is the BLUP of $\boldsymbol{\alpha}$ from both sources.

3 Multiplicative Benchmarking Models

The survey error series, $\{u_1(t)\}$, in the first source of an additive model [the more frequent one, i.e., (1.1)], usually can be written as

$$u_1(t) = \sigma(t)\tilde{u}(t), \tag{3.1}$$

where $\sigma(t)$ is the standard deviation of $u_1(t)$ and $\{\tilde{u}(t)\}$ is a stationary series of mean zero and variance one with either the autocorrelation function or a model for it known. In practice, $\sigma(t)$ estimated from a survey is not constant. Although heteroscedasticity apparently does not affect the application of signal extraction formulae like (2.15), it causes problems in modelling $\{\alpha(t)\}$: it is very difficult to estimate V_a from the data. Sometimes the heteroscedasticity is due to the estimation error in survey process. Bell and Hillmer (1990) suggested that when the fluctuation of these estimated values of $\sigma(t)$ is small, $\sigma(t)$ may be regarded as a constant and they recommended to use the average of the estimates of $\sigma(t)$ for practical purposes.

However, often heteroscedasticity is intrinsic in the sense that $\sigma(t)$ depends strongly on the level of $\alpha(t)$. In such case, we can put

$$u_1(t) = \alpha(t)c(t)\tilde{u}_1(t), \tag{3.2}$$

where $c(t) = \sigma(t)/\alpha(t)$ is the CV. Then, the observations can be written in a multiplicative form as (1.2):

$$y_1(t) = \alpha(t)(1 + c(t)\tilde{u}_1(t)), \tag{3.3}$$

where $c(t)\tilde{u}_1(t)$, the series of percentage errors, is independent of $\{\alpha(t)\}$.

For most socio-economic series, $y_1(t)$ and $\alpha(t)$ are positive. Taking logarithm in (3.3) and denoting

$$y_*(t) = \log y_1(t), \quad \alpha_*(t) = \log \alpha(t), \quad u_*(t) = \log(1 + c(t)\tilde{u}(t)), \tag{3.4}$$

we obtain an linear model

$$y_*(t) = \alpha_*(t) + u_*(t). \tag{3.5}$$

Usually, the distribution of the logarithm transformed series can be assumed normal. Under this assumption, (3.5) is a standard additive model for signal extraction. Let t run from 1 to n. Then we may put (3.5) in vector form as $\mathbf{y}_* = \boldsymbol{\alpha}_* + \mathbf{u}_*$. If we assume that V_{u_*}, V_{α_*} and $E\boldsymbol{\alpha}_*$ are known, then (2.15) may be used to obtain the extracted signal and the variance matrix of errors [correspondences: $\mathbf{y}_1 \leftrightarrow \mathbf{y}_*$, $\boldsymbol{\alpha} \leftrightarrow \boldsymbol{\alpha}_*$, $\mathbf{u}_1 \leftrightarrow \mathbf{u}_*$, $\hat{\boldsymbol{\alpha}}^{(1)}(t) \leftrightarrow \hat{\boldsymbol{\alpha}}_*$, and $V_{\hat{e}^{(1)}} \leftrightarrow V_*$]. i.e., we obtain

$$\hat{\boldsymbol{\alpha}}_* := E(\boldsymbol{\alpha}_* \mid \mathbf{y}_*), \quad V_* := E\{(\hat{\boldsymbol{\alpha}}_* - \boldsymbol{\alpha}_*)(\hat{\boldsymbol{\alpha}}_* - \boldsymbol{\alpha}_*)'\}. \tag{3.6}$$

Since $\boldsymbol{\alpha}_*$, $\mathbf{y}_*$ are jointly normal, V_* given by (3.6) is also the conditional variance, i.e. $V_* = \mathrm{Var}\,(\boldsymbol{\alpha}_* \mid \mathbf{y}_*)$. In other words, given $\mathbf{y}_*$, the conditional distribution of $\boldsymbol{\alpha}_*$ is $N(\hat{\boldsymbol{\alpha}}_*, V_*)$, then given $\mathbf{y}_1$ (equivalent to "given $\mathbf{y}_*$"), the conditional distribution of $\boldsymbol{\alpha} = (e^{\alpha_*(1)} \ldots e^{\alpha_*(n)})'$ is lognormal and from the established result (see Johnson and Kotz, 1972, p.20), we have

$$\hat{\alpha}^{(1)}(t) := E\{\alpha(t) \mid \mathbf{y}_1\} = \exp\{\hat{\alpha}_*(t) + v_*(t)/2\}, \tag{3.7}$$

$$\begin{aligned}\mathrm{Cov}\,\{\alpha(s), \alpha(t) \mid \mathbf{y}_1\} &= E\{(\hat{\alpha}^{(1)}(s) - \alpha(s))(\hat{\alpha}^{(1)}(t) - \alpha(t)) \mid \mathbf{y}_1\} \\ &= (\exp\{V_*(s,t)\} - 1)\exp\{\hat{\alpha}_*(s) + \hat{\alpha}_*(t) + (v_*(s) + v_*(t))/2\},\end{aligned} \tag{3.8}$$

where $V_*(s,t)$ is the (s,t)th entry of V_* and $v_*(t) = V_*(t,t)$ which do not depend on $\mathbf{y}_1$; however $\hat{\alpha}_*(s)$ and $\hat{\alpha}_*(t)$ do depend on $\mathbf{y}_1$. In the following, the meaning of $V_\alpha(s,t), v_\alpha(t)$, etc, are the similar.

Thus, we have the BUP of $\boldsymbol{\alpha}$ from the first source, given by (3.7), which can be used as the initial value of the recursive procedure for obtaining the BFUP. However, to implement recursive formulae (2.5) and (2.6), we also need

$$V_{\hat{e}^{(1)}} := \mathrm{Var}\,(\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha}) = E\{E[(\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha})' \mid \mathbf{y}_1]\} = E\{\mathrm{Var}\,(\boldsymbol{\alpha} \mid \mathbf{y}_1)\}, \quad (3.9)$$

where the (s,t)th entry of $\mathrm{Var}\,(\boldsymbol{\alpha} \mid \mathbf{y}_1)$ is given by (3.8). Hence, for obtaining $V_{\hat{e}^{(1)}}$, we have to calculate $E(e^{\hat{\alpha}_*(s)}e^{\hat{\alpha}_*(t)})$. Putting $\boldsymbol{\alpha}_* - E\boldsymbol{\alpha}_* = (\boldsymbol{\alpha}_* - \hat{\boldsymbol{\alpha}}_*) + (\hat{\boldsymbol{\alpha}}_* - E\boldsymbol{\alpha}_*)$ and observing that $E\{(\boldsymbol{\alpha}_* - \hat{\boldsymbol{\alpha}}_*)\,(\hat{\boldsymbol{\alpha}}_* - E\boldsymbol{\alpha}_*)'\} = 0$, we have

$$V_{\hat{\alpha}_*} := \mathrm{Var}\,(\hat{\boldsymbol{\alpha}}_*) = V_{\alpha_*} - V_*. \quad (3.10)$$

Since $\hat{\boldsymbol{\alpha}}_*$ defined by (3.6) is also normal with mean zero and variance matrix $V_{\hat{\alpha}_*}$, the random vector $(e^{\hat{\alpha}_*(1)} \ \ldots \ e^{\hat{\alpha}_*(n)})'$ is lognormal. By the same established result mentioned above and noticing $E\hat{\alpha}_*(t) = E\alpha_*(t)$, we have

$$E(e^{\hat{\alpha}_*(s)}e^{\hat{\alpha}_*(t)}) = \exp\{E\alpha_*(s) + E\alpha_*(t) + (v_{\hat{\alpha}_*}(s) + v_{\hat{\alpha}_*}(t))/2 + V_{\hat{\alpha}_*}(s,t)\}. \quad (3.11)$$

Finally, (3.8) through (3.11) lead to

$$V_{\hat{e}^{(1)}}(s,t) = (\exp\{V_*(s,t)\} - 1)\times$$

$$\exp\{E\alpha_*(s) + E\alpha_*(t) + (v_{\alpha_*}(s) + v_{\alpha_*}(t))/2 + V_{\alpha_*}(s,t) - V_*(s,t)\}. \quad (3.12)$$

According to the above discussion, for the model given by (1.2) and (1.3), the procedure to obtain the BFUP of $\boldsymbol{\alpha}$ from $\mathbf{y}_1$ and $\mathbf{y}_2$ (the benchmarked values) can be summarized in 4 steps as follows.

1. Take logarithm as (3.4). Assume $\mathbf{u}_* \sim N(0, V_{u_*})$, where $\mathbf{u}_*$ is approximated by (3.13), below, with known $c(t)$ and covariance function of $\{\tilde{u}(t)\}$, so V_{u_*} is known.

2. Assume $\boldsymbol{\alpha}_*$ is normal and independent of $\mathbf{u}_*$. Use signal extraction formulae, (2.15), with indicated correspondence above (3.6) to obtain $\hat{\boldsymbol{\alpha}}_*$ and V_* in (3.6), where $E\boldsymbol{\alpha}_*$ and V_{α_*} are assumed to be known.

3. Using obtained $\hat{\boldsymbol{\alpha}}_*$ and V_* as well as the known values of $E\boldsymbol{\alpha}_*$ and V_{α_*}, (3.7) and (3.12) give $\hat{\boldsymbol{\alpha}}^{(1)}$, the BUP of $\boldsymbol{\alpha}$ from $\mathbf{y}_1$, and $V_{\hat{e}^{(1)}}$, the covariance matrix of its prediction error.

4. Placing the obtained values of $\hat{\boldsymbol{\alpha}}^{(1)}$ and $V_{\hat{e}^{(1)}}$ in (2.5) and (2.6) (in one step or recursively if the benchmarks, $\mathbf{y}_2$, can be regarded as from more than one source).

Since $c(t) > 0$ in (3.4) is small, so approximately,

$$u_*(t) = c(t)\tilde{u}(t). \tag{3.13}$$

The autocovariance of $\tilde{u}(t)$ can be given by survey experts. As far as $c(t)$ is concerned, in fact, statistical agencies usually publish CV rather than $\sigma(t)$, so these $c(t)$ are known (estimates are obtained during the survey process), and then, we may have V_{u_*}. Similar to what was mentioned at the beginning of this section, if the fluctuation of $c(t)$ is small, we would suggest using their average c to replace $c(t)$. In fact, most of the statistical agencies design their surveys in order to have $c(t) \simeq c$ for all t, where c is a small positive constant; so, this condition usually fulfills.

A problem was posed in the implementation of Step 2, as well as of step 3: For obtaining $\hat{\boldsymbol{\alpha}}_*$ and V_*, except $V_{u_*}^{-1}$, we also need $V_{\alpha_*}^{-1}$ and $E\boldsymbol{\alpha}_*$ [again, cf. (2.15)], and in (3.12), V_{α_*} and $E\boldsymbol{\alpha}_*$ are directly used. But all these are usually unknown.

As we have mentioned at the end of Remark 2.4, given V_{u_*}, the estimation of V_{α_*} and $E\boldsymbol{\alpha}_*$ from data, $\mathbf{y}_*(t)$, has been discussed by Chen *et al* (1997), where $\alpha_*(t)$ can be modelled as a TS or a DS series. A remark is that, in this case, the TS model is preferable rather than the DS model chosen to model $\alpha_*(t)$, because V_{α_*} is not unique under the DS model assumption. Although, we may choose $V_{\alpha_*}^{-1}$ [corresponding to V_α^{-1} in (2.15)] resulting from the diffuse assumption for the early values of $\alpha_*(t)$ which can lead to very nice results for predicting $\alpha_*(t)$ (signal extraction), however, as we also directly apply V_{α_*} in (3.12), the diffuse assumption is improper for V_{α_*}. Moreover, in (3.12), we need $E\alpha_*(t)$ which is not easy to be estimated from data under the DS model assumption because the mean is hidden polynomials and seasonality and the series is also nonstationary in variance. [This mean is eliminated in (2.15), i.e. $V_\alpha^{-1}E\boldsymbol{\alpha} = 0$, under the DS model assumption, as it was pointed out by Chen *et al*, 1997, around (22).]

Appendix Proof of Theorem 2.1

(1) Put $\mathbf{y}_2 - X_2\hat{\boldsymbol{\alpha}}^{(1)} = -X_2\hat{\mathbf{e}}^{(1)} + \mathbf{u}_2$. Since $\mathbf{u}_2$ is uncorrelated with $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}^{(1)}$, so it is uncorrelated with $\hat{\mathbf{e}}^{(1)}$. Then from (2.5) one can verify that $V_{\hat{e}^{(2)}} = \mathrm{Var}\,(\hat{\boldsymbol{\alpha}}^{(2)} - \boldsymbol{\alpha})$ is given by (2.6). Note that (2.3) and (2.4) imply $X_2 V_{\hat{e}^{(2)}} X_2' + V_{u_2} > 0$. The same formulae as (2.5) and (2.6) hold for $\tilde{\boldsymbol{\alpha}}^{(2)}$ and $V_{\tilde{e}^{(2)}}$.

Let $\delta > 0$,

$$\hat{\mathbf{y}}_2 = \begin{pmatrix} \hat{\boldsymbol{\alpha}}^{(1)} \\ \mathbf{y}_2 \end{pmatrix}, \quad \dot{X}_2 = \begin{pmatrix} I_n \\ X_2 \end{pmatrix}, \quad \hat{\mathbf{u}}_2 = \begin{pmatrix} \hat{\mathbf{e}}^{(1)} \\ \mathbf{u}_2 \end{pmatrix}, \tag{A.1}$$

and let the block diagonal matrix

$$V_{\hat{u}_2}(\delta) = \mathrm{diag}(V_{\hat{e}^{(1)}}(\delta),\ V_{u_2}(\delta)) := \mathrm{diag}(V_{\hat{e}^{(1)}} + \delta I_n,\ V_{u_2} + \delta I_{n_2}) \tag{A.2}$$

be the variance matrix of $\hat{\mathbf{u}}_2$. Formally using the variance formula of GLS for model $\hat{\mathbf{y}}_2 = \dot{X}_2\boldsymbol{\alpha} + \hat{\mathbf{u}}_2$, i.e. letting

$$V_{\hat{e}^{(2)}}(\delta) := [\dot{X}_2' V_{\hat{u}_2}(\delta)\dot{X}_2]^{-1} = [V_{\hat{e}^{(1)}}(\delta)^{-1} + X_2' V_{u_2}(\delta)^{-1} X_2]^{-1}, \tag{A.3}$$

the well-known identity from the partitioned inverse of a matrix leads to

$$V_{\hat{e}^{(2)}}(\delta) = V_{\hat{e}^{(1)}}(\delta) - V_{\hat{e}^{(1)}}(\delta)X_2'[X_2 V_{\hat{e}^{(1)}}(\delta)X_2' + V_{u_2}(\delta)]^{-1} X_2 V_{\hat{e}^{(1)}}(\delta). \tag{A.4}$$

Similarly, let $V_{\tilde{u}_2}(\delta) = \mathrm{diag}(V_{\tilde{e}^{(1)}}(\delta),\ V_{u_2}(\delta))$ we have $V_{\tilde{e}^{(2)}}(\delta)$ defined as in (A.3) which can be expressed as in (A.4) with $V_{\tilde{e}^{(1)}}(\delta)$ replacing $V_{\hat{e}^{(1)}}(\delta)$. From (2.3), $V_{\hat{u}_2}(\delta) \le V_{\tilde{u}_2}(\delta)$ holds. Then (A.3) gives

$$V_{\hat{e}^{(2)}}(\delta) \le V_{\tilde{e}_{(2)}}(\delta). \tag{A.5}$$

Let $\delta \to 0$, then $V_{\hat{e}^{(1)}}(\delta) \to V_{\hat{e}^{(1)}}$ and $V_{u_2}(\delta) \to V_{u_2}$ and hence by (A.4) and (2.6), $V_{\hat{e}^{(2)}}(\delta) \to V_{\hat{e}^{(2)}}$. Similarly $V_{\tilde{e}^{(2)}}(\delta) \to V_{\tilde{e}^{(2)}}$. Then (2.7) follows from (A.5).

(2) First, we show that (2.9) is equivalent to (2.4) and (2.10). Put the left side of (2.9) as

$$\begin{pmatrix} X_2 V_{\hat{e}^{(1)}} X_2' + V_{u_2} & X_2 V_{\hat{e}^{(1)}} X_3' \\ X_3 V_{\hat{e}^{(1)}} X_2' & X_3 V_{\hat{e}^{(1)}} X_3' + V_{u_3} \end{pmatrix} := \begin{pmatrix} B_{22} & B_{23} \\ B_{32} & B_{33} \end{pmatrix} := B. \tag{A.6}$$

Then (2.4) implies $B_{22} > 0$. Let

$$B^* = \begin{pmatrix} I_{n_2} & 0 \\ -B_{32}B_{22}^{-1} & I_{n_3} \end{pmatrix} B \begin{pmatrix} I_{n_2} & -B_{22}^{-1}B_{23} \\ 0 & I_{n_3} \end{pmatrix} = \begin{pmatrix} B_{22}^{-1} & 0 \\ 0 & B_{33} - B_{32}B_{22}^{-1}B_{23} \end{pmatrix}; \tag{A.7}$$

from (2.6) and (A.6), it is easy to verify that $B_{33} - B_{32}B_{22}^{-1}B_{23} = X_3 V_{\hat{e}^{(2)}} X_3'$, so both (2.4) and (2.10) hold if and only if $B^* > 0$, or $B > 0$, or (2.9) holds.

Keep the notations in (2.8) and let $\hat{\mathbf{y}}_3$, $\dot{X}_3$ and $\hat{\mathbf{u}}_3$ be similar to (A.1), (by adding $\mathbf{y}_3$, X_3 and $\mathbf{u}_3$ to the bottoms of $\hat{\mathbf{y}}_2$ and so on, correspondingly). Similarly, let

$$V_{\hat{u}_3}(\delta) = \text{diag}(V_{\hat{e}^{(1)}}(\delta), V_{\bar{u}_3}(\delta)), \qquad V_{\bar{u}_3}(\delta) = \text{diag}(V_{u_2}(\delta), V_{u_3}(\delta)) \tag{A.8}$$

be the covariance matrix of $\hat{\mathbf{u}}_3$ and $\bar{\mathbf{u}}_3$; let

$$\begin{aligned} V_{\hat{e}^{(3)}}(\delta) &:= [\dot{X}_3' V_{\hat{u}_3}(\delta) \dot{X}_3]^{-1} = [V_{\hat{e}^{(1)}}(\delta)^{-1} + \bar{X}_3' V_{\bar{u}_3}(\delta)^{-1} \bar{X}_3]^{-1} \\ &= [V_{\hat{e}^{(2)}}(\delta)^{-1} + X_3' V_{u_3}(\delta)^{-1} X_3]^{-1}. \end{aligned} \tag{A.9}$$

From the middle expression of $V_{\hat{e}^{(3)}}(\delta)$ in (A.9) and due to the similar identity which links (A.3) and (A.4), we get the first expression of (2.13) by letting $\delta \to 0$. Applying the same procedure to the last expression of (A.9), we see that its limit is the last expression of (2.13). As both expressions in (2.13) are the limit of $V_{\hat{e}^{(3)}}(\delta)$, they are identical.

Formally, using the estimation formula of GLS for model $\hat{\mathbf{y}}_3 = \dot{X}_3 \boldsymbol{\alpha} + \hat{\mathbf{u}}_3$ with $V_{\hat{u}_3}(\delta)$ as the variance matrix of $\hat{\mathbf{u}}_3$ and defining

$$\begin{aligned} \hat{\boldsymbol{\alpha}}^{(3)}(\delta) &:= V_{\hat{e}^{(3)}}(\delta) \dot{X}_3 V_{\hat{u}_3}(\delta)^{-1} \hat{\mathbf{y}}_3 \\ &= V_{\hat{e}^{(3)}}(\delta) V_{\hat{e}^{(1)}}(\delta)^{-1} \hat{\boldsymbol{\alpha}}^{(1)} + V_{\hat{e}^{(3)}}(\delta) \bar{X}_3' V_{\bar{u}_3}(\delta)^{-1} \bar{\mathbf{y}}_3, \end{aligned} \tag{A.10}$$

then the middle expression of $V_{\hat{e}^{(3)}}(\delta)$ in (A.9) and the similar identity which links (A.3) and (A.4) lead to

$$\begin{aligned} &\hat{\boldsymbol{\alpha}}^{(3)}(\delta) = \hat{\boldsymbol{\alpha}}^{(1)} - V_{\hat{e}^{(1)}}(\delta) \bar{X}_3' [\bar{X}_3 V_{\hat{e}^{(1)}}(\delta) \bar{X}_3' + V_{\bar{u}_3}(\delta)]^{-1} \bar{X}_3 \hat{\boldsymbol{\alpha}}^{(1)} + \\ &V_{\hat{e}^{(1)}}(\delta) \bar{X}_3' \{I_n - [\bar{X}_3 V_{\hat{e}^{(1)}}(\delta) \bar{X}_3' + V_{\bar{u}_3}(\delta)]^{-1} \bar{X}_3 V_{\hat{e}^{(1)}}(\delta) \bar{X}_3'\} V_{\bar{u}_3}(\delta)^{-1} \bar{\mathbf{y}}_3. \end{aligned} \tag{A.11}$$

The last term in (A.11) can be simplified to $V_{\hat{e}^{(1)}}(\delta) \bar{X}_3' \, [\bar{X}_3 V_{\hat{e}^{(1)}}(\delta) \bar{X}_3' + V_{\bar{u}_3}(\delta)]^{-1} \bar{\mathbf{y}}_3$. Let $\delta \to 0$, $\hat{\boldsymbol{\alpha}}^{(3)}(\delta) \to \hat{\boldsymbol{\alpha}}^{(3)}$ be defined as (2.11).

Let $\hat{\boldsymbol{\alpha}}^{(2)}(\delta) := V_{\hat{e}^{(2)}}(\delta) \dot{X}_2 \, V_{\hat{u}_2}(\delta)^{-1} \hat{\mathbf{y}}_2$. When $\delta \to 0$, by the same discussion as above, we have $\hat{\boldsymbol{\alpha}}^{(2)}(\delta) \to \hat{\boldsymbol{\alpha}}^{(2)}$ defined as (2.5). Expressing $\hat{\boldsymbol{\alpha}}^{(3)}(\delta)$ defined in (A.10) alternatively by

$$\hat{\boldsymbol{\alpha}}^{(3)}(\delta) = V_{\hat{e}^{(3)}}(\delta) V_{\hat{e}^{(2)}}(\delta)^{-1} \hat{\boldsymbol{\alpha}}^{(2)}(\delta) + V_{\hat{e}^{(3)}}(\delta) X_3' V_{u_3}(\delta) \mathbf{y}_3, \tag{A.12}$$

and at this time using the last expression of $V_{\hat{e}^{(3)}}(\delta)$ in (A.9), the same discussion concludes that, when $\delta \to 0$, $\hat{\boldsymbol{\alpha}}^{(3)}(\delta)$ converges to (2.12). So, (2.12) and (2.11) are identical.

(3) The property of the BLUP can be proved by using the formulae of projection (see Chen *et al*, 1996; Durbin and Quenneville, 1997). However it may also be proved by a similar method given below in (4) by defining $\mathcal{A} = \{\boldsymbol{\alpha} : \boldsymbol{\alpha} = A_1\mathbf{y}_1 + A_2\mathbf{y}_2\}$, where A_i is any $n \times n_i$ matrix of constants ($i = 1, 2$). Obviously, $\hat{\boldsymbol{\alpha}}^{(2)}$ is a FUP.

(4) Since $\hat{\boldsymbol{\alpha}}^{(1)}$ is a vector of measurable functions of $\mathbf{y}_1$, so $\hat{\boldsymbol{\alpha}}^{(2)}$ given by (2.5) is in $\mathcal{A}$ defined by (2.14). For proving the minimization of $V_{\hat{e}^{(2)}}$, it is sufficient to show that

$$E[(\hat{\boldsymbol{\alpha}}^{(2)} - \boldsymbol{\alpha})\check{\boldsymbol{\alpha}}'] = 0 \quad \text{for any} \quad \check{\boldsymbol{\alpha}} \in \mathcal{A} \tag{A.13}$$

due to the following reason. Temporarily denote $\hat{\boldsymbol{\alpha}}^{(2)}$ by $\hat{\boldsymbol{\alpha}}$ and $(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^2 := (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})'$ etc. For any $\tilde{\boldsymbol{\alpha}} \in \mathcal{A}$, let $\check{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}$. Obviously, $\check{\boldsymbol{\alpha}} \in \mathcal{A}$. Now

$$E(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^2 = E(\tilde{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}})^2 + E(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^2 + E[\check{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})'] + E[(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\check{\boldsymbol{\alpha}}']. \tag{A.14}$$

The cross terms disappear due to (A.13) and hence, for all $\tilde{\boldsymbol{\alpha}} \in \mathcal{A}$, (A.14) is minimized at $\tilde{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}$.

From (2.5),

$$\hat{\boldsymbol{\alpha}}^{(2)} - \boldsymbol{\alpha} = [I_n - V_{\hat{e}^{(1)}}X_2'(X_2V_{\hat{e}^{(1)}}X_2' + V_{u_2})^{-1}X_2](\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha})$$

$$+V_{\hat{e}^{(1)}}X_2'(X_2V_{\hat{e}^{(1)}}X_2' + V_{u_2})^{-1}\mathbf{u}_2. \tag{A.15}$$

Since $\hat{\boldsymbol{\alpha}}^{(1)} = E(\boldsymbol{\alpha} \mid \mathbf{y}_1)$, so for any vector of measurable functions $\check{\mathbf{a}}(\mathbf{y}_1)$, $E[(\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha})\ \check{\mathbf{a}}(\mathbf{y}_1)'] = 0$. Since $E\mathbf{u}_2 = 0$ and $\mathbf{u}_2$ is independent of $\mathbf{y}_1$, so $E[\mathbf{u}_2\ \check{\mathbf{a}}(\mathbf{y}_1)'] = 0$. Then from (A.15) we have

$$E[(\hat{\boldsymbol{\alpha}}^{(2)} - \boldsymbol{\alpha})\ \check{\mathbf{a}}(\mathbf{y}_1)'] = 0. \tag{A.16}$$

On the other hand, again due to the fact that $\mathbf{u}_2$ is independent of $\mathbf{y}_1$ and $\boldsymbol{\alpha}$, and hence independent of $\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha}$, we have

$$E(\mathbf{u}_2\mathbf{y}_2') = V_{u_2}, \qquad E[(\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha})\mathbf{u}_2'] = 0, \tag{A.17}$$

Then

$$E[(\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha})\mathbf{y}_2'] = E[(\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha})\boldsymbol{\alpha}']X_2' = -V_{\hat{e}^{(1)}}X_2'. \qquad (A.18)$$

From (A.15), (A.17) and (A.18), making use of the formulae from partitioned inverse, we may check that

$$E[(\hat{\boldsymbol{\alpha}}^{(2)} - \boldsymbol{\alpha})\mathbf{y}_2'] = 0. \qquad (A.19)$$

Then (A.13) does hold due to (A.16), (A.19) and (2.14). That completes the proof of the theorem.

Acknowledgment

We are indebted to P.Cholette for his most valuable comments. Some results in this paper are further developments from our earlier works jointly with him. We are grateful to D.Binder, J.Durbin, I.MacNeill, B.Quenneville and A.Singh for stimulating discussions and their very important suggestions.

References

Bell, W.R. and Hillmer, S.C.(1990). The time series approach to estimation for repeated survey. *Survey Methodology*, **16**, 195-215.

Chambers, J.M. (1971). Regression updating. *Journal of American Statistical Association*, **66**, 744-748.

Chen, Z.-G., Cholette, P.A. and Dagum, E.B. (1996). A nonparametric method for benchmarking survey data via signal extraction. Working Paper No. BSMD 96-001E, Statistics Canada.

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

Chen, Z.-G. and Dagum, E.B. (1998). Conditions and procedures for optimal prediction of variables with constraints from multiple sources. Time Series Research and Analysis Centre, Statistics Canada.

Chen, Z.-G. and Ni, J.-Y. (1989). Subset regression time series and its modelling procedures. *Journal of Multivariate Analysis*, **31**, 266-288.

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. Forthcoming *International Statistical Review*

Dempfle, L. (1977). Comparison of several sire evaluation methods in dairy cattle breeding. *Livestock Production Science*, **4**, 129-139.

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

Godambe, V.P. (1960). An optimum property of regular maximum likelihood estimation. *The annals of Mathematical Statistics*, **31**, 8-12.

Godambe, V.P. and Thompson, M.E. (1989). An extension of quasi-likelihood estimation (with discussion). *Journal of Statistical Planning and Information*, **22**, 137-172.

Goldberger, A.S. (1962). Best linear unbiased prediction in generalized linear regression model. *Journal of American Statistical Association*, **57**, 369-375.

Harville, D.A. (1990). BLUP (best linear unbiased prediction) and Beyond. In *Advances in Statistical Methods in genetic improvement of livestock* (D. Geanola and K. Hammond, eds.), 239-276. Springer, New York.

Henderson, C.R.(1950). Estimation of genetic parameters (abstract). *Annals of Mathematical Statistics*, **21**, 309-310.

Henderson, C.R.(1975). Best linear unbiased estimation and prediction under a selection model *Biometrics*, **31**, 423-447.

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

Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate*

Distribution, John Wiley & Sons, Inc. New York, London, Sydney, Toronto.

Jones,R.H. (1970). Recursive estimation of subset of regression coefficients. *The Annals of Mathematical Statistics*, **41**, 688-691.

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

Odell, P.L. and Lewis, T.O. (1971). Best linear recursion estimation. *Journal of American Statistical Association.* **66**, 893-896.

Robinson, G.K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science*, **6**, 15-51.

Singh, A.C. (1995). Predicting functions for generalization of BLUP to mixed nonlinear models. *Proceedings of American Statistical Association*, Biometrics Section, 300-305.

Singh, A.C. and Kovacevic, M.S. (1996). Multidimensional benchmarking of time series by segmented Kalman filtering. *Proceedings of American Statistical Association*, Business and Economic Statistics Section, 40-45.

Weale, M.R.(1992). Estimation of data measured with error and subject to linear restrictions. *Journal of Applied Economics.* **7**, 167-174.

Whittle, P. (1963). *Prediction and Regulation by Linear Least-Square Methods.* University of Minnesota Press. Minneapolis, Minnesota.