Statistics Canada Statistique Canada

Methodology Branch

Household Survey
Methods Division

Direction de la méthodologie

Division des méthodes
d'enquêtes-ménages

Canada

MULTIDIMENSIONAL BENCHMARKING OF TIME SERIES BY SEGMENTED KALMAN FILTERING

HSMD - 96 - 004E

A.C. Singh[1,2] and M.S. Kovacevic[1]

[1]Survey and Analysis Methods Development Section, HSMD, Statistics Canada, 16-A, R.H. Coats, Ottawa, Ontario K1A 0T6

[2]Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S 5B6

Acknowledgement The paper was presented in a session sponsored by the Business and Economics Statistics Section at the Joint Statistical Meetings in Chicago, August 4-8, 1996. The first author's research was supported in part by a grant from Natural Sciences and Engineering Research Council of Canada.

MULTIDIMENSIONAL BENCHMARKING OF TIME SERIES BY SEGMENTED KALMAN FILTERING

ABSTRACT

Benchmarking is essentially a method of signal estimation from time series under constraints. The final signal estimates satisfy the regression-adjusted benchmarks in the nonbinding case, but are forced to exactly satisfy benchmarks in the binding case; the latter case leads to sub-optimality in the case of random benchmarks. It is assumed that the source of benchmark series is independent of the source of target time series. Typically, the process of benchmarking consists of two stages: the first stage for finding initial signal estimates and the second stage for constrained regression. When the number of benchmarks is quite large as in the case of multidimensional benchmarking, the usual method of constrained regression may be computationally difficult due to high dimension of matrix inversion involved therein. If benchmarks are independent of each other, then the technique of recursive least squares can be adapted to avoid matrix inversion. For dependent benchmarks, a method termed Segmented Kalman Filtering (SKF) is proposed which alleviates the above computational difficulty under very general conditions. Moreover, with SKF, it is generally easy to revise later parts of the target time series in light of new benchmarks without revising the earlier parts and without losing optimality, when the new benchmarks do not involve signals from earlier parts. This feature may be quite convenient for implementation in practice as it will not require computer manipulations of the full MSE matrix of signal estimates; the dimension of this matrix may be quite large for the multidimensional problem or for a long target series. In the interest of simplicity, the proposed method is illustrated using a simulated bivariate time series from a random walk plus noise model. It is believed that SKF provides an important statistical tool which may have wide applications.

Key Words: Covariance norm; Gram-Schmidt orthogonalization; Recursive least squares; Zero functions

ÉTALONNAGE MULTIDIMENSIONNEL DES SÉRIES CHRONOLOGIQUES PAR FILTRAGE DE KALMAN SEGMENTÉ

RÉSUMÉ

L'étalonnage est essentiellement une méthode d'estimation du signal à partir des séries chronologiques sous contraintes. Dans le cas des données repères non fermes, les estimations finales du signal satisfont les données repères corrigées par régression; pour les données repères fermes, il doit toutefois y avoir satisfaction exacte, ce qui entraîne une sous-optimalité des données repères aléatoires. On présume que la source de la série de référence est indépendante de la source de la série chronologique visée. En général, le processus d'étalonnage comporte deux étapes : d'abord, estimations initiales du signal puis, régression assujettie à une contrainte. Lorsque le nombre de données repères est assez élevé, comme c'est le cas avec l'étalonnage multidimensionnel, la méthode habituelle de régression assujettie à une contrainte peut être difficile à réaliser, au point de vue du calcul, en raison de l'ordre élevé d'inversion de matrice. Si les données repères sont indépendantes les unes des autres, la technique des moindres carrés récursifs peut alors être adaptée pour éviter l'inversion de matrice. Pour les données repères dépendantes, la méthode dite filtrage de Kalman segmenté (FKS) est proposée; celle-ci réduit les problèmes de calcul précités dans des conditions très générales. De plus, avec la méthode FKS, il est généralement facile de réviser les parties ultérieures de la série chronologique visée à la lumière des nouvelles données repères, sans avoir à réviser les parties antérieures et sans perdre d'optimalité, lorsque les nouvelles données repères ne font pas intervenir de signaux des parties antérieures. Cette particularité pourrait se révéler fort utile en pratique, car il ne sera pas nécessaire de procéder à des manipulations informatiques de la matrice complète de l'erreur quadratique moyenne des estimations des signaux; l'ordre de cette matrice peut être assez élevé lorsqu'il s'agit de problèmes multidimensionnels ou d'une longue série chronologique. Par souci de simplicité, la méthode proposée est illustrée à l'aide d'une série chronologique simulée à deux variables, obtenue par processus de marche aléatoire avec modélisation du bruit. La méthode FKS apparaît comme un important outil statistique qui pourrait avoir de nombreuses applications.

Mots clés : norme de covariance, orthogonalisation de Gram-Schmidt, moindres carrés récursifs, fonctions zéro

1. INTRODUCTION

The problem considered in this article is how to use information in an auxiliary time series of benchmarks to revise a target time series $\{ y_t ,\ 1 \le t \le T\}$ in order to get more precise estimates of signal parameters. The signal parameters may be random or nonrandom; random signals may be serially independent or dependent. Typically, serially independent signals will be nonrandom in practice. For an example of nonrandom signals, consider a time series where the observation y_t at time t represents the signal parameter η_t except for contamination with noise ϵ_t and possibly bias b_t, i.e., the expectation μ_t of y_t, is equal to $\eta_t + b_t$. . Thus the nonrandom signals are defined as expectation of y_t after adjustment for bias if any. On the other hand, for random signal parameters, suppose μ_t is a function of parameters θ_t. The θ_t parameters include, in general, both fixed and random components, and some or all of the random components evolve over time. Now the random signal parameters η_t can be defined as functions of θ_t-parameters. The benchmark series, $\{x_s\!:1 \le s \le S\}$, provides auxiliary information about signals $\{\eta_t\}$ in the sense that x_s is a linear function of signals plus random error. If the benchmarks x_s are nonrandom, then there is no random error, and thus they are necessarily binding. However, if x_s are random, they could be either binding or nonbinding. The case of binding but random benchmarks arises in practice when the benchmark series is considered sufficiently reliable and therefore, no smoothing is warranted. In this case the final signal estimates are forced to exactly satisfy the benchmarks even if they are random. However, in the nonbinding case, they satisfy the regression-adjusted benchmarks, after taking into account the sampling error in benchmarks.

There may be several types of benchmark series. For example, if $\{ y_t \}$ is a univariate monthly series, one may have annual benchmark x-series, and quarterly benchmark z-series from auxiliary surveys; here both x- and z- series represent two types of temporal aggregation constraints. If $\{ y_t \}$ is a multivariate monthly series, one may have annual multivariate benchmark x-series, and monthly univariate benchmark z-series which provides information about the aggregate of components for each month; here the x-series represents temporal benchmarks across time for each component time series while the z-series represents contemporaneous benchmarks across component for each time. Thus, we can classify the benchmarking problem as either unidimensional or multidimensional. In the unidimensional case, there is only one type of benchmark constraints, whereas in the multidimensional case, there are two or more types.

For simplicity, we consider only two types of benchmark series $\{x_s; 1 \le s \le S\}$ and $\{z_r; 1 \le r \le R\}$. Suppose, we extend the original series $\{y_t\}$ by augmenting at the end the benchmark series. Thus, the extended series will consist of the y-series segment followed by the x-series segment, which in turn is followed by the z-series segment. The order of benchmark segments is arbitrary but, in practice, may depend on the availability of benchmark information. Now the problem of revising y-series via benchmarking can be thought of as a huge linear regression problem which will involve, in general, both fixed and random parameters. We can perform (stochastic) least squares to get "optimal" (in the sense of BLUE or BLUP, the best linear unbiased estimation or prediction, as the case may be) estimates of signals such that they satisfy either regression-adjusted or the original benchmarks depending on the nature of benchmarks. If the benchmarks are random but binding, the resulting signal estimates will only be suboptimal.

The interest in benchmarking has a long history; for an early reference, see Stone et al.(1942). For the problem under consideration, the main papers are among others due to Denton (1971), Hillmer and Trabelsi (1987), and Cholette and Dagum (1994). The paper of Hillmer and Trabelsi represents a milestone in the benchmarking literature in that unlike the traditional numerical approach it develops a proper statistical framework. For the unidimensional benchmarking problem under the assumption that the source of benchmark information is independent of the source of the target y-series, they proposed a two-stage procedure: (i) initial signal estimation under a model which treats signal parameters either as serially dependent (e.g., in the case of ARIMA or structural modelling) or independent (e.g., in the case of seasonal adjustment via X11-ARIMA modelling), (ii) adjustment (or smoothing) of initial signal estimates to satisfy benchmarks via constrained regression. Section 2 provides a review of existing benchmarking methods.

Although the existing methods can be adapted to deal with the multidimensional problem, there are mainly two concerns regarding the computational complexity that may arise: (i) The matrix inversion involved in each of the two stages may be computationally prohibitive (especially in the case of multidimensional benchmarking) in view of its high dimension, the dimension being the number of signal parameters for the first stage, and the number of benchmarks for the second stage, (ii) With a very long target series $\{y_t\}$ or with many component series in the multidimensional case, size of the mean square matrix (MSE) of signal estimates may be too large for computer manipulations. Section 3 gives a heuristic motivation for addressing these two concerns.

We propose a method termed Segmented Kalman Filtering (SKF) which makes an attempt to alleviate the concerns raised above for existing methods, and is applicable under general conditions. Section 4 contains theoretical considerations underlying the proposed method of SKF. The SKF method is described in Section 5, and as a simple illustration, a numerical example based on simulated data from a bivariate random walk plus noise model is given in Section 6. The final Section 7 contains concluding remarks.

2. REVIEW OF EXISTING METHODS

As mentioned in the introduction, the existing methods consider only the unidimensional benchmarking problem and use, in general, a two stage solution to the benchmarking problem: (i) Initial signal estimation, to be denoted by $\{\tilde{\eta}_t\}$, and (ii) Adjusting $\{\tilde{\eta}_t\}$ to get the final signal estimates, denoted by $\{\hat{\eta}_t\}$, such that the benchmarks are satisfied. Now given $\{\tilde{\eta}_t\}$, the benchmarking problem is simply that of constrained regression. We have

$$\tilde{\eta} = \eta + \delta, \quad \delta \sim (0, \Omega) \tag{2.1a}$$

$$x = L\eta + e, \quad e \sim (0, \Sigma_e), \tag{2.1b}$$

where δ and e are orthogonal in the sense of being uncorrelated. The optimal (in the sense of BLUE) $\hat{\eta}$ is given by

$$\hat{\eta} = \tilde{\eta} + \Omega L'(L\Omega L' + \Sigma_e)^{-1}(x - L\tilde{\eta}) \tag{2.2}$$

If the benchmarks are nonrandom, then $\Sigma_e = O$. If they are random but binding, then the (suboptimal) $\hat{\eta}$ is obtained by setting Σ_e to O. However, for the random but nonbinding case, Σ_e can not be set to O, and the benchmarks are not exactly satisfied. The amount of adjustment in $L\tilde{\eta}$ depends on its MSE $L\Omega L'$ relative to the MSE Σ_e of benchmarks. In other words, the benchmarked signal estimates $\{\hat{\eta}\}$ satisfy the regression-adjusted benchmarks and not the original benchmarks.

Now, depending on the nature of the initial signal estimation, the benchmarking problem can be classified into two types.

2.1 Serially Independent Signals

Here, the model for the initial signal estimation treats signals as serially independent; nonrandom in particular. For example, using the X11-ARIMA method of seasonal adjustment, the signal series may be defined as the expectation of the seasonally adjusted y-series. The corresponding MSE matrix Ω may be obtained approximately using (sampling) design or modelling considerations. Then the benchmarking problem essentially reduces to that of constrained regression as mentioned above. This sort of regression approach was taken by Cholette and Dagum (1994) who also allowed for bias and autocorrelation in the survey estimates represented by the y-series. There is an alternative but equivalent way of viewing this constrained regression approach in terms of a minimization problem where the distance, $(\tilde{\eta} - \eta)'\Omega^{-1}(\tilde{\eta} - \eta)$, is minimized subject to benchmark constraints. A practical interpretation of this is that the initial estimate series $\{\tilde{\eta}_t\}$ is perturbed only a little to satisfy the benchmarks while preserving relevant characteristics of the series. This is somewhat similar to the traditional method of Denton (1971) which is a numerical procedure based on distance minimization. Cholette and Dagum made an important observation that the commonly used Denton-type method can be obtained as a special case of the regression approach using a suitable working MSE matrix for Ω. In contrast to the above semiparametric approach, one can also use a parametric approach based on, for example the method of maximum likelihood, to get $\hat{\eta}_t$ from $\tilde{\eta}_t$, see, e.g., Mian and Laniel (1993) under the assumptions of normality and constant multiplicative bias.

2.2 Serially Dependent Signals

The assumption of dependent signals is useful when the series itself does not represent signals, but contains them as random unobserved components (e.g., trend, seasonal etc.). The objective is to estimate signal components under benchmark constraints. This will, in general, give rise to more efficient and smoother signal estimates than those for the independent signal case.

Often ARIMA or basic structural modelling (BSM) are used for modelling $\{y_t\}$. Hillmer and Trabelsi (1987) used ARIMA to estimate $\{\tilde{\eta}_t\}$ and to specify the corresponding MSE matrix Ω. As shown by Bell (1984), this can be rendered into a problem of signal estimation via state space modelling. This gives the initial signal estimates for the first stage, which are adjusted in the second stage via constrained regression to satisfy benchmarks. Alternatively, Durbin and Quenneville (1996) used BSM to estimate $\{\tilde{\eta}_t\}$ and their MSE matrix. They allowed for more general nonstationary time series, heteroscedasticity in survey errors, bias in survey estimates, and the nonlinear case where the basic model is multiplicative in components but the benchmarks are additive in components; for a discussion, see Singh (1995a). Another alternative was proposed by Chen, Cholette and Dagum (1995) which uses a non-parametric approach for estimating initial signals and their MSE matrix.

3. HEURISTIC MOTIVATION

In this section we address the two main concerns regarding the computational complexity of the existing benchmarking methods as stated in the introduction.

For the first concern, namely, that of the possibility of high dimensionality in matrix inversion at each of the two stages, consider first the problem at the second stage. For this problem it seems natural to propose processing of constrained regression into substages via subset regression on benchmarks; in particular, processing benchmarks piecewise, i.e., one at a time. The basic idea is to transform the auxiliary information in benchmarks into zero functions (i.e., functions which are zero in expectation), and then to orthogonalize them (e.g., by the Gram-Schmidt method) in the sense that they become uncorrelated. Note that at any given substage of constrained regression, the benchmarked signal estimates are simply regression estimates, i.e., are obtained as a residual after regressing (or projecting) on the subset of predictors (in the form of zero functions constructed from benchmarks). It follows that the adjusted signal estimates at each substage satisfy the subset of (regression-adjusted) benchmarks involved at that stage and that they continue to satisfy the benchmarks used in earlier substages due to the orthogonalization of the corresponding zero functions. This important property, namely, that the benchmark-adjusted signal estimates at any substage do not disturb the benchmarks used in earlier substages, provided benchmarks are suitably orthogonalized, is key to the proposed method of piecewise processing of benchmarks.

For the orthogonalization of benchmark information mentioned above, it will be convenient in practice if it can be implemented recursively, i.e., using only estimates of the model parameters (denoted by θ_t) for the current time t to orthogonalize new benchmarks with respect to the old ones. Note that a recursive processing is necessarily piecewise, but not vice-versa. If benchmarks are serially independent, then one can adapt the technique of recursive least squares. However, for the general case of dependent benchmarks, one can borrow ideas from the Kalman Filter (KF) orthogonalization in the state space framework, see also Odell and Lewis (1971) for a related discussion. In particular, the two conditions, independence (or orthogonality) of new benchmarks at time t of the old ones conditional on the θ_t-parameters, and the Markovian dependence of future θ-parameters on the current and past parameters, are required. These two conditions are not very restrictive and are met by commonly used time series models. Both conditions can often be realized in practice by suitably enlarging the parameter (or the state) vector. Thus, it would be possible to use a KF-type algorithm on segments of the benchmark series, e.g., first on the x-segment and then on the z-segment.

The problem of high dimension in matrix inversion at the first stage (due to a large number of signal parameters) can also be dealt with in a similar manner, i.e., by using a state space framework for the target y-series, and then a KF for obtaining initial signal estimates.

For addressing the second concern, we consider the possibility of revising later parts of the time series in light of new benchmarks without revising the earlier parts and without losing optimality, when the new benchmarks do not involve signals from earlier parts. Having this feature will not only reduce the size of the MSE matrix of signal estimates, but also will be quite convenient from practical considerations. The reason for this is that when the new y-series and benchmark information become available, one will have the flexibility of deciding whether or not the full y-series should be revised. Moreover, the signal estimates corresponding to the earlier parts of the y-series will generally be not affected much by the new benchmarks. If the y-series can also be cast in the state space framework in addition to the benchmark series, then one does not lose optimality of the revised later part of the series when the earlier part is not revised in light of new benchmarks. Notice that the state space framework of y-series generally implies that the signal parameters are random and that they evolve over time. However, it can also encompass independent or nonrandom but time-varying signal parameters as a limiting case by letting the variance of the corresponding error in the transition equation go to infinity.

In the following, we assume that for a time period of interest (to be denoted by $1 \leq t \leq T$), the y-series augmented with benchmarks x- and z- series is given. In particular, this may contain only the later part of the original y-series and only the new benchmark series as discussed above.

4. PIECEWISE AND RECURSIVE PROCESSING: THEORETICAL CONSIDERATIONS

In this section, we consider conditions for piecewise and recursive processing of information in target and benchmark series using the theory of zero functions (see, Rao, 1968). Note that zero functions (also known as elementary estimating functions) are simply functions which are zero in expectation, and are, in general, functions of both data and parameters. It will be seen that converting new pieces of information in target and benchmark series into parameter-free orthogonal zero functions gives rise to innovations which, in turn, can be conveniently used for piecewise updating of estimates. It is known for special types of regression that a recursive (and hence piecewise) procedure can be applied. For example, in the case of independent observations, recursive least squares can be used, and in the case of dependent observations cast in the state space framework, the Kalman Filter (KF) provides the recursive processing of data to get optimal estimates.

For the general regression problem we shall first consider conditions required for piecewise processing of information in data.

4.1 Piecewise Processing for General Regression

Consider a general regression problem with a p-vector of fixed parameters β, for $i = 1, \ldots, n$

$$y_i = x_i' \beta + \epsilon_i, \tag{4.1}$$

where the n-vector $\epsilon \sim (\mathbf{0}, \mathbf{\Gamma})$, and is uncorrelated with the $n \times p$ matrix X of observations on the predictors. The least squares estimate, $\hat{\beta} = (X' \Gamma^{-1} X)^{-1} X' \Gamma^{-1} y$, can be obtained alternatively by piecewise regression on parameter-free orthogonal zero functions created from the y-observations. Letting $\hat{\beta}^{(1)}$ denote the initial estimate of β based on the first p y-values, (i.e., $\hat{\beta}^{(1)} = X_p^{-1} y_p$, say), it follows that $\hat{\beta}$ can be obtained by regressing $\hat{\beta}^{(1)}$ on the $(n-p)$ zero functions $g_1 = y_{p+1} - x_{p+1}' \hat{\beta}^{(1)}, \ldots, g_{(n-p)} = y_n - x_n' \hat{\beta}^{(1)}$, that is, as a residual after projecting on the $(n-p)$ predictors (in the form of zero functions) under the covariance norm. The functions $g_1, \ldots, g_{n-p}$ are parameter-free, and can be orthogonalized via Gram-Schmidt to get innovations $g_1^*, \ldots g_{n-p}^*$. Now $\hat{\beta}$ can be obtained in a piecewise manner as follows:

$$\begin{aligned}
\hat{\beta}^{(2)} &= \hat{\beta}^{(1)} - Cov(\hat{\beta}^{(1)}, g_1^*)[V(g_1^*)]^{-1} g_1^*, \\
\hat{\beta}^{(3)} &= \hat{\beta}^{(2)} - Cov(\hat{\beta}^{(2)}, g_2^*)[V(g_2^*)]^{-1} g_2^*, \\
&\vdots
\end{aligned} \tag{4.2}$$

and so on until $\hat{\beta}^{(n-p+1)}$ which equals $\hat{\beta}$.

We remark that if β contains some random parameters, then for the corresponding stochastic least squares version (see, for example, Singh, 1995a), an analogous piecewise procedure can be easily defined. We also remark that in the context of our extended time series for the benchmarking problem, the above piecewise processing of information in the benchmark segment to get signal estimates $\{\hat{\eta}_t\}$ from the initial estimates $\{\tilde{\eta}_t\}$ (which correspond to $\hat{\beta}^{(1)}$) will not disturb the benchmark constraints processed earlier on. This follows from the basic principle of regression estimation (viewed as an orthogonal projection on predictor zero functions under the covariance norm) that the regression estimate of a predictor zero function itself is exactly zero.

Next we consider under what conditions, the above piecewise procedure for finding $\hat{\beta}$ can be made recursive, i.e., for the regression problem (4.1), letting t denote the current time, we want to create an innovation from y_t by using only the most recent estimate $\hat{\beta}^{(t-1)}$ (representing a condensed form of past data) and not the full past data $y_1, \ldots, y_{t-1}$. The resulting innovation will, in turn, give $\hat{\beta}^{(t)}$. This is similar to the KF algorithm for state space models where although β, in general, varies with t but is connected over time through a Markovian relation to allow recursive updating. In the next subsection, we consider a sufficient set of conditions provided by the state space regression model for recursive processing of information. This is accomplished through a brief introduction to KF via zero functions. A simple state space framework in terms of random walk plus noise model will be used for this purpose.

4.2 Recursive Processing for State Space Regression Models

Suppose the regression model is specified in terms of two equations (measurement and transition equations of the state space setup) for the random walk plus noise model as follows: for $1 \leq t \leq T$,

$$y_t = \theta_t + \epsilon_t, \qquad \epsilon_t \sim WN(0, \sigma_\epsilon^2) \tag{4.3a}$$

$$\theta_t = \theta_{t-1} + \xi_t, \qquad \xi_t \sim WN(0, \sigma_\xi^2), \tag{4.3b}$$

where WN signifies white noise, ϵ_t's and ξ_t's are orthogonal and ξ_t is also orthogonal to θ_{t-1}. The first equation (4.3a) implies the independence (or orthogonality) of new information from the past conditional on the (random) state parameter θ_t, and the second equation (4.3b) implies Markovian dependence between the θ-parameters. Note that the θ-parameters are random and vary with t. Alternatively, the usual single equation representation of the regression model (4.3) in terms of the current parameter θ_T, is given by, for $1 \le t \le T$,

$$y_t = \theta_T - \sum_{t'=t+1}^{T} \xi_{t'} + \epsilon_t, \tag{4.4}$$

where the corresponding covariance structure can be appropriately specified. For other parameters, θ_t, similar equations can be defined.

Now, at $t = 1$, the best estimate (i.e., BLUP) $\hat{\theta}_{1|1}$ of θ_1 is y_1 because we have only one piece of information in the form of zero function $g_1 = y_1 - \theta_1$. At $t = 2$, we get another zero function $y_2 - \theta_2$. It can be made parameter-free by replacing θ_2 by $\hat{\theta}_{2|1} = y_1$, where $\hat{\theta}_{2|1}$ denotes an estimate of θ_2 based on information available at t=1. Therefore $g_2 = y_2 - y_1$. Similarly, we have $g_3 = y_3 - y_1, \ldots, g_T = y_T - y_1$ as we get more information. We will show that because of the state space framework, $\{g_t : 2 \le t \le T\}$ can be orthogonalized recursively to get innovations.

At time t=1, the BLUP of θ_2 is the sum of the BLUPs of θ_1 and ξ_2, i.e. $\hat{\theta}_{2|1} = \hat{\theta}_{1|1} + \hat{\xi}_{2|1}$. In view of the Markovian relation (4.3b), the error ξ_2 does not depend on the past θ_1 and therefore, $\hat{\xi}_{2|1} = 0$, implying $\hat{\theta}_{2|1} = \hat{\theta}_{1|1}$. Next consider $t = 2$. The first innovation from y_2, denoted by g_2^* can be set equal to $g_2 = y_2 - \hat{\theta}_{2|1} = y_2 - y_1$. Now, the BLUP of θ_2 combines optimally two pieces of information, $\hat{\theta}_{2|1}$ and $g_2^* = y_2 - \hat{\theta}_{2|1}$, available at t=2, and is given by

$$\hat{\theta}_{2|2} = \hat{\theta}_{2|1} - Cov(\hat{\theta}_{2|1} - \theta_2, y_2 - \hat{\theta}_{2|1})\,[V(y_2 - \hat{\theta}_{2|1})]^{-1}(y_2 - \hat{\theta}_{2|1}), \tag{4.5a}$$

$$= \hat{\theta}_{2|1} + V(\hat{\theta}_{2|1} - \theta_2)\,[\sigma_\epsilon^2 + V(\hat{\theta}_{2|1} - \theta_2)]^{-1}\,(y_2 - \hat{\theta}_{2|1}). \tag{4.5b}$$

Next, we need to orthogonalize the zero function g_3 ($= y_3 - y_1$) with g_2^* to get the innovation g_3^* from y_3. First note that $\hat{\theta}_{3|2} = \hat{\theta}_{2|2}$ because $\hat{\xi}_{3|2} = 0$ again using the Markovian relation (4.3b). Notice that $\hat{\theta}_{3|2}$ is obtained in a recursive manner. Second, in view of the conditional independence of ϵ_3 from the past ϵ_2 and ϵ_1, the innovation g_3^* obtained from y_3 is $g_3^* = y_3 - \hat{\theta}_{3|2} = y_3 - \hat{\theta}_{2|2}$. Similarly, we can get innovations from the zero functions $y_t - y_1$, $t = 4,\ldots,T$, as

$$g_t^* = y_t - \hat{\theta}_{t|t-1} = y_t - \hat{\theta}_{t-1|t-1} \tag{4.6}$$

and, in the process, get optimal estimates $\hat{\theta}_{2|2},\ldots,\hat{\theta}_{T|T}$ recursively as

$$\hat{\theta}_{t|t} = \hat{\theta}_{t|t-1} + V(\hat{\theta}_{t|t-1} - \theta_t)\,[\sigma_\epsilon^2 + V(\hat{\theta}_{t|t-1} - \theta_t)]^{-1}\,(y_t - \hat{\theta}_{t|t-1}),\; t = 2,\ldots,T. \tag{4.7}$$

The equation (4.7) is known as the filtering equation of KF. The variance term $V(\hat{\theta}_{t|t-1} - \theta_t)$ is also computed in a recursive manner. The final estimate $\hat{\theta}_{T|T}$ is termed the smoothed estimate of θ_T based on all the y-observations. Similarly, smoothed estimates $\hat{\theta}_{t|T}$ for $t = 1,\ldots,T-1$ can be obtained by regressing $\hat{\theta}_{t|t}$ (already computed) on the innovations $g_{t+1}^*,\ldots,g_T^*$.

It can be seen from the above brief introduction of KF in terms of a simple state space model that for recursive processing of the benchmarks in the second stage, we need to cast each benchmark segment into the state space framework. In other words, we need the two conditions: the independence of benchmarks conditional on state parameters (suitably chosen), and Markovian dependence between the state parameters over time. We can then create innovations (via orthogonalization) from the benchmarks using the signal estimates $\tilde{\eta}$ obtained from y-series only. Note that for obtaining the signal estimates $\tilde{\eta}$, information in the target y-series can be processed recursively via KF if y-series also follows a state space model. Now, the desired (benchmarked) signal estimates $\hat{\eta}$ are automatically obtained after running KF over benchmark segments of the extended time series where $\tilde{\eta}$ is used for the initialization of KF. This is the basic principle underlying the proposed method of segmented KF given in the next section. Note that for the piecewise recursive processing via a state space model it is assumed that the underlying covariance structure is suitably specified from the sampling design and model (either true or working) considerations.

5. SEGMENTED KALMAN FILTER: The Proposed Method

Consider the (multivariate) target series $\{y_t\colon\ 1 \le t \le T\}$ and the two benchmark series $\{x_s\colon\ 1 \le s \le S\}$ and $\{z_r\colon\ 1 \le r \le R\}$. Suppose all the three segments of the extended time series can be cast in the state space framework as follows (cf: Harvey, 1989, ch. 3):

<u>y-segment</u> $(1 \le t \le T)$

$$y_t = F_t^y \theta_t^y + \epsilon_t^y, \quad \epsilon_t^y \sim (0, \Gamma_t^y) \tag{5.1a}$$

$$\theta_t^y = G_t^y \theta_{t-1}^y + \xi_t^y, \quad \xi_t^y \sim (0, \Lambda_t^y) \tag{5.1b}$$

<u>x-segment</u> $(1 \le s \le S)$

$$x_s = F_s^x \theta_s^x + \epsilon_s^x, \quad \epsilon_s^x \sim (0, \Gamma_s^x) \tag{5.2a}$$

$$\theta_s^x = G_s^x \theta_{s-1}^x + \xi_s^x, \quad \xi_s^x \sim (0, \Lambda_s^x) \tag{5.2b}$$

<u>z-segment</u> $(1 \le r \le R)$

$$z_r = F_r^z \theta_r^z + \epsilon_r^z, \quad \epsilon_r^z \sim (0, \Gamma_r^z) \tag{5.3a}$$

$$\theta_r^z = G_r^z \theta_{r-1}^z + \xi_r^z, \quad \xi_r^z \sim (0, \Lambda_r^z) \tag{5.3b}$$

The model errors ϵ's and ξ's are assumed to satisfy the usual orthogonality conditions. The signal of interest η_t is a linear function of components of the state vector θ_t^y for $1 \le t \le T$, and therefore, its estimate $\hat{\eta}_{t|T}$ and the corresponding MSE matrix can be obtained from those for θ^y-parameters. Also, the complete signal vector $\eta = (\eta_1', \ldots, \eta_T')'$ over the T time points is assumed to be a subset of the state vector θ_s^x for each s and θ_r^z for each r. Thus, signal estimates $\hat{\eta}_{t|T,S}$, $\hat{\eta}_{t|T,S,R}$ and their MSE matrices can be obtained respectively from those of θ_s^x and θ_r^z parameters. The notation $\hat{\eta}_{t|T,S,R}$, for example, signifies smoothed (or BLUP) estimates of η_t after information in T y-observations, S x-benchmarks, and R z-benchmarks is utilized.

The proposed method of SKF consists of the following steps. Each step corresponds to processing a segment of the extended time series.

Step I (KF for the y-segment)

This step is without benchmarks and one can use the usual KF (and smoother) to get smoothed estimates of $\{\theta_t^y: 1 \leq t \leq T\}$ and the corresponding MSE matrix P^y. These, in turn give the signal estimates $\{\hat{\eta}_{t|T}\}$, to be denoted by $\hat{\eta}^{(1)}$, and their MSE matrix $\Omega^{(1)}$. Note that if the y-series is not cast in the state space set-up, then this step of KF will not be needed. Instead, the signal estimates and their MSE matrix are obtained either through a general regression model or a working model based on subject-matter considerations.

Step II (KF for the x-segment)

This KF is different from the usual KF in that $\hat{\eta}^{(1)}$ and $\Omega^{(1)}$ obtained from Step I are used for initialization of that part of the θ_s^x- vector which contains the η-parameters. If there are some additional parameters they may be initialized in the usual way. Then, the KF is run as usual to get the estimate $\hat{\theta}_{S|S}^x$ and its MSE $P_{S|S}^x$. This gives rise to $\hat{\eta}^{(2)}$ and $\Omega^{(2)}$. Note that unlike Step I, there is no need of Kalman smoothing because the complete η-vector is contained in each θ_s^x by construction and therefore $\hat{\eta}^{(2)} = \{\hat{\eta}_{t|T,S}\}$ and its MSE matrix $\Omega^{(2)}$ can be obtained from $\hat{\theta}_{S|S}^x$ and $P_{S|S}^x$. Also, note that if the x-benchmarks are independent (which may be true in the case of annual x-benchmarks for monthly y-series), then θ_s^x will be simply the complete signal vector η in which case it does not evolve over s, i.e., the transition equation (5.2b) becomes the trivial one, namely, $\theta_s^x = \theta_{s-1}^x$, with zero error.

Step III (KF for the z-segment)

This is similar to Step II except that $(\hat{\eta}^{(2)}, \Omega^{(2)})$ is used for initialization of the η- part of the state vector θ_r^z. The required signal estimates $\hat{\eta}^{(3)} = \{\hat{\eta}_{t|T,S,R}\}$ and its MSE matrix $\Omega^{(3)}$ can be obtained from $\hat{\theta}_{R|R}^z$ and $P_{R|R}^z$.

Remark 5.1 If more y-data are expected in future, as is usually the case, then it is better to include all θ^y-parameters, and not just the signal parameters, in each of θ_s^x and θ_r^z parameters.

This way when the new observation series becomes available, one can use the current estimate of θ_T^y as input for KF initialization for the next step (IV). Similarly, for the new benchmarks, we will have Steps V, VI, and so on. The final signal estimates can always be obtained from the final θ^y-estimates.

Remark 5.2 If both x- and z-benchmarks are treated as binding, then clearly they must be made compatible with each other. For instance, for a multidimensional problem if x-benchmarks are annual and z- benchmarks are monthly, then the annual totals of z-benchmarks should agree with x-benchmarks. Also, whenever random benchmarks are treated as fixed, the MSE matrix of the resulting signal estimates should be adjusted for variability of the random benchmarks. This can be done along the lines of Pfeffermann and Burck (1990).

Remark 5.3 The above procedure for SKF assumes that the hyperparameters (i.e., parameters of design matrices $\boldsymbol{F}$, transition matrices $\boldsymbol{G}$, and MSE matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$) are given in advance. In practice, they are usually estimated by MLE under normality. The log-likelihood can be easily modified to include auxiliary information from benchmarks via additional innovations (or orthogonalized zero functions) obtained from benchmarks as follows:

$$\log\text{-}likelihood = const - (1/2)[\textstyle\sum_t \log v_t^y + \sum_s \log v_s^x + \sum_r \log v_r^z]$$

$$- (1/2)[\textstyle\sum_t (g_t^{*y})^2/v_t^y + \sum_s (g_s^{*x})^2/v_s^x + \sum_r (g_r^{*z})^2/v_r^z]$$

where g^*'s and v's represent innovations and their MSE respectively for y-, x- and z-segments.

Remark 5.4 To implement SKF in practice, suitable modelling of the target and benchmark series is required. In the context of survey sampling, if the observations and benchmarks represent design-consistent estimates, then a design-based version of SKF (denote by d-SKF) can be easily defined under a working model to obtain design-consistent benchmarked estimates provided all the zero functions appearing in benchmarked estimates are design-consistent estimates of zero. The resulting estimates are robust in that they remain design-consistent even if the working model assumptions are not correct. (However, finding a suitable estimate of the corresponding MSE matrix may not be easy in general.) The working model may be chosen, for instance, to correspond to the Denton-type distance minimization method (see the example in Subsection 6.1.1) such that important characteristics (growth rates, for example) of the target series are approximately preserved. An important application of d-SKF may be in the area of benchmarking seasonally adjusted component series in a multidimensional problem where both the contemporaneous and temporal benchmarks are treated as binding, assuming of course their compatibility. The contemporaneous benchmarks, in practice, can be obtained by first benchmarking (via d-SKF) the seasonally-adjusted aggregate series (over components) with respect to the aggregate temporal benchmarks, thus ensuring their compatibility with temporal benchmarks. This type of hierarchical approach for ensuring compatibility may be particularly useful when there are many dimensions in the benchmarking problem.

Remark 5.5 It should be emphasized, as mentioned in the introduction, that the order of processing of benchmark segments is arbitrarily chosen.

6. ILLUSTRATION

For illustrating SKF, we simulated a five year long bivariate monthly series following a random walk plus noise model given by, for $j = 1,2$

$$y_{jt} = \theta^y_{jt} + \epsilon^y_{jt}, \qquad \epsilon^y_{jt} \sim_{iid} N(0, \sigma^{2(y)}_{j\epsilon}) \tag{6.1a}$$

$$\theta^y_{jt} = \theta^y_{j,t-1} + \xi^y_{jt}, \qquad \xi^y_{jt} \sim_{iid} N(0, \sigma^{2(y)}_{j\xi}). \tag{6.1b}$$

The signal η_{jt}, in general, is a function of θ^y_{jt}; in this example η_{jt} equals θ^y_{jt}. The two components y_{1t} and y_{2t} were made correlated by introducing the correlation $\rho(\epsilon^y_{1t}, \epsilon^y_{2t})$ which was set as .25. The values of $\sigma^{2(y)}_{1\epsilon}$, $\sigma^{2(y)}_{2\epsilon}$, $\sigma^{2(y)}_{1\xi}$ and $\sigma^{2(y)}_{2\xi}$ were set at 4×10^3, 16×10^3, $.625\times10^3$ and 2.5×10^3 respectively which give rise to signal-noise variance ratio as .156 for each j. The starting values for θ^y_{jt} were set at 500 and 800 for $j = 1,2$ respectively. For each component j, the five annual benchmarks were obtained from the generated (true) signals η_{jt}, thus making the benchmarks nonrandom, $x_{js} = \sum_{t=12(s-1)}^{12s} \theta^y_{jt}$, for

$s = 1,\ldots,5$.

The monthly benchmarks were allowed to be either serially independent or dependent. For the independent case, the aggregates of (true) monthly signals over the two components were used as monthly benchmarks, $z_r = \theta^y_{1r} + \theta^y_{2r}$, for $r = 1,\ldots,60$. Thus these monthly benchmarks become nonrandom. For the dependent case, a noise series following AR(1) with $\rho = .8$ and $\sigma^2 = 100$ was added to the monthly aggregates of signals; the last 60 values out of a total of 300 generated values for the noise series were used in the interest of achieving stationarity. The resulting benchmarks were treated as nonbinding for the example considered. Tables 1 and 2 show the generated y-series $\{y_{jt}\}$ for the first two years as well as the annual $\{x_{js}\}$ and monthly benchmarks $\{z_{jr}\}$.

6.1 Nonrandom Annual and Monthly Benchmarks

6.1.1. Serially Independent Signals

Although the true signals are serially dependent, we first treat them as nonrandom and use the limiting case of state space modelling (when the error variance in the transition equation tends to infinity) for signal estimation for the first step of SKF. The underlying regression model is equivalent to a Denton-type method corresponding to minimization of the distance function

$$\Sigma_j \, \Sigma_t \, [(\Delta\eta_{jt} - \Delta y_{jt})/y_{j,t-1}\, \sigma_{j\delta}]^2 \text{ subject to benchmarks,} \qquad (6.2)$$

where $\eta_{jt} = \theta^y_{jt}$, $\Delta\eta_{jt}$ and Δy_{jt} denote first differences, and $\sigma^2_{j\delta}$ provides differential weights for the distance over different components $j = 1,2$; these weights do not appear for the unidimensional problem. The above minimization problem approximately preserves the rate of growth in the y-series after benchmarking.

The regression model is chosen in a manner somewhat similar to the one suggested by Cholette and Dagum (1994), with the working covariance structure corresponding to (6.2):

$$
\begin{aligned}
&\text{at } t = 1, \quad && y_{j1} = \eta_{j1} + \epsilon_{j1}, \quad \epsilon_{j1} \sim (0, \sigma_{j\epsilon}^2), \\
&\text{for } t \geq 2, \quad && \Delta y_{jt} = \Delta \eta_{jt} + \delta_{jt}, \quad \delta_{jt} \sim (0, \sigma_{j\delta}^2 y_{j,t-1}^2),
\end{aligned}
\tag{6.3}
$$

with $\sigma_{j\epsilon}^2$ tending to infinity, implying that the first observation is discarded in estimating the signals. However, first differences of the nonrandom signals are estimated by the corresponding first differences of the y-series, that is

$$
\Delta \hat{\eta}_{jt} = y_{jt} - y_{j,t-1}, \quad t \geq 2. \tag{6.4}
$$

The signal estimate for $t=1$ $\hat{\eta}_{j1}$ is obtained after the first annual benchmark (x_{j1}) becomes available. The reason for this is that x_{j1} can be represented as

$$
x_{j1} = \sum_{t=1}^{12} \eta_{jt} = \eta_{j1} + \sum_{t=2}^{12} (12 - t + 1)\, \Delta \eta_{jt}, \tag{6.5}
$$

which implies

$$
\hat{\eta}_{j1} = [x_{j1} - \sum_{t=2}^{12} (12 - t + 1)\, \Delta \hat{\eta}_{jt}] / 12. \tag{6.6}
$$

Using (6.4), (6.6) and the covariance structure given by (6.3) we obtain the signal estimates $\hat{\eta}^{(1)} = \{\hat{\eta}_{jt}\}$ and their MSE matrix $\Omega^{(1)}$. This completes Step I of SKF except that variance components $\sigma_{j\delta}^2$ need to be specified. They are estimated from the observation and benchmark series by a simple method of moments (not shown here) as .3006 and .3096 respectively for $j = 1,2$.

Now, for higher steps of SKF, we need to specify hyperparameters of state space models for the benchmark segments. However, since all the benchmarks are nonrandom, state space models for benchmark series become trivial as in the case of recursive least squares, i.e., errors in the transition equations (5.2b) and (5.3b) become zero so that:

$$
\theta_s^x = \theta_{s-1}^x, \quad s = 1, \ldots, S \tag{6.7}
$$

$$
\theta_r^z = \theta_{r-1}^z, \quad r = 1, \ldots, R \tag{6.8}
$$

where $\theta_s^x = \theta_r^z = (\eta_{11}, \eta_{21}; \eta_{12}, \eta_{22}; \eta_{13}, \eta_{23}; \ldots; \eta_{1T}, \eta_{2T})'$. For higher steps of SKF we consider the four scenarios regarding the availability of benchmarks: (i) annual benchmarks for the first year, (ii) annual and monthly benchmarks for the first year, (iii) annual benchmarks for the first two years but monthly only for the first year, (iv) annual and monthly benchmarks for the first two years.

Step II of SKF, in this example, will coincide with Step I because the first annual benchmark was already used in Step I. However, in general, this is not the case and Step II can be described as follows when the annual benchmark x_{j1} becomes available. The observation equation (5.2a) is $x_s = F_s^x \theta_s^x$ with the matrix $F_s^x = \left[O_{1s} \mid H^x \mid O_{2s}\right]$ with O_{1s} and O_{2s} defined as zero matrices of order $2 \times 24(s-1)$ and $2 \times (2T-24s)$ respectively, and H^x denotes a 2×24 matrix

$$H^x = \begin{bmatrix} 1 & 0 & 1 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 1 \end{bmatrix}.$$

The initial estimate $\hat{\theta}_0^x$ is obtained from Step I as $\hat{\theta}_0^x = \hat{\eta}^{(1)} = (\hat{\eta}_{11}, \hat{\eta}_{21}, \ldots, \hat{\eta}_{1T}, \hat{\eta}_{2T})'$ with MSE $P_0^x = \Omega^{(1)}$. The BLUP, $\hat{\theta}_{1|0}^x$ of θ_1^x is $\hat{\theta}_0^x$ with the MSE $P_{1|0}^x = P_0^x$. When the benchmark x_1 becomes available, the innovation g_1^{x*}, at $s=1$, is $x_1 - F_1^x \hat{\theta}_{1|0}^x$, with MSE $G_1 = E(g_1^{x*} g_1^{x*\prime}) = F_1^x P_{1|0}^x F_1^{x\prime}$. Now the BLUP of θ_1^x is

$$\hat{\theta}_{1|1}^x = \hat{\theta}_0^x + P_{1|0}^x F_1^{x\prime} G_1^{-1} g_1^{x*} = \hat{\theta}_0^x + K_1^x g_1^{x*} \tag{6.9}$$

with MSE $P_1^x = (I - K_1^x F_1^x) P_{1|0}^x$. This completes Step II which gives rise to $\hat{\eta}^{(2)}$ and $\Omega^{(2)}$.

For step III of SKF the observation equation (5.3a) takes the form $z_r = F_r^z \theta_r^z$ with matrix F_r^z partitioned as $F_r^z = [\mathbf{0}_{1r}' \mid H^{z\prime} \mid \mathbf{0}_{2r}']$, where $\mathbf{0}_{1r}$ and $\mathbf{0}_{2r}$ are zero vectors of length $2(r-1)$ and $2(T-r)$ respectively, and $H^z = (1, 1)'$. The initial estimate $\hat{\theta}_0^z$ is obtained from Step II as $\hat{\theta}_0^z = \hat{\eta}^{(2)}$ with MSE $P_0^z = \Omega^{(2)}$. The BLUP, $\hat{\theta}_{1|0}^z$ of θ_1^z is $\hat{\theta}_0^z$ with the MSE $P_{1|0}^z = P_0^z$. Once the benchmark z_1 becomes available, the innovation g_1^{z*}, at $r=1$, is $z_1 - F_1^z \hat{\theta}_{1|0}^z$, with MSE $G_1 = E(g_1^{z*} g_1^{z*\prime}) = F_1^z P_{1|0}^z F_1^{z\prime}$. Then the BLUP of θ_1^z is

$$\hat{\theta}^z_{1|1} = \hat{\theta}^z_0 + P^z_{1|0} F^{z\,\prime}_1 \, G^{-1}_1 \, g^{z\,*}_1 = \hat{\theta}^z_0 + K^z_1 \, g^{z\,*}_1 \tag{6.9}$$

with MSE $P^z_1 = (I - K^z_1 F^z_1) P^z_{1|0}$. After utilization of the first twelve monthly benchmarks $z_1,\ldots,z_{12}$ we obtain $\hat{\theta}^z_{R|R}$ and $P^z_{R|R}$ for $R=12$. This completes Step III and gives $\hat{\eta}^{(3)}$ and $\Omega^{(3)}$.

Table 1(a) shows the revised y-series (denoted by $\hat{\eta}^{(1)}_t$, $\hat{\eta}^{(2)}_t$, $\hat{\eta}^{(3)}_t$, $\hat{\eta}^{(4)}_t$, $\hat{\eta}^{(5)}_t$) for the first two years under the four scenarios. Observe that the earlier parts of the series are generally not affected much by benchmarks involving signals corresponding to later parts of the series. Also note that the second component gets more affected. This is as expected due to the high variability of the error δ_{2t} relative to δ_{1t} in the model (6.3).

6.1.2 Serially Dependent Signals

In this case, for using a state space model for the initial signal estimation for Step I of SKF, we chose the same random walk plus noise model that was used for generating the y-series with the same hyperparameters for illustration purposes. For initialization of KF, the observation at t=1 for each component was used. The next steps of SKF are similar to those used in section 6.1.1. Table 1(b) shows the revised y-series as the benchmarks become available under the four scenarios. Notice that the series looks smoother than that for nonrandom signals. Also observe that the earlier parts of the series are generally not affected much in light of the new benchmarks.

6.2 Nonrandom Annual and Serially Dependent but Nonbinding Monthly Benchmarks

6.2.1 Serially Independent Signals

Here the Step I of SKF is same as that for Subsection 6.1.1. However, in Step II for the segment of monthly benchmarks for the first year, we need to specify a state space model because of serially dependent benchmarks. For the sake of illustration, we chose the same model that was used for generating these benchmarks and the same hyperparameters. Now KF for this segment is performed, and then in Step III for the first annual benchmark, only a trivial state space model is needed for SKF, and similarly for other steps. Table 2(a) shows the revised y-series as the benchmarks become available under the four scenarios. Notice that as expected the monthly aggregate series does not match the monthly nonbinding benchmarks.

6.2.2 Serially Dependent Signals

This is similar to the previous case except that the initial signals are estimated using a state space model as in Subsection 6.1.2. The benchmarked y-series in this case is shown in Table 2(b).

Figures 1 and 2 corresponding to tables 1(a) and 2(b) respectively show graphs of revised target series for the first two year period as benchmarks become available under different scenarios corresponding to serially independent and dependent monthly benchmarks, the annual benchmarks being nonrandom in all cases. Figures for tables 1(b) and 2(a) are not shown here but are similar to those for tables 1(a) and 2(b) respectively.

7. CONCLUDING REMARKS

By approaching Kalman Filtering from the perspective of orthogonal zero functions, the method of segmented Kalman Filtering for revising the target series was proposed which can process benchmarks piecewise and recursively under the conditions of state space modelling for the benchmark series. The SKF makes a single pass through each benchmark. The final signal estimates satisfy all the benchmarks (which are regression-adjusted in the nonbinding case). The SKF method also allows for revising only later parts of y-series in light of new benchmarks without losing optimality, again under state space modelling assumptions for the y-series. The proposed method encompasses Denton-type numerical methods based on distance minimization by using a suitable working covariance matrix for the initial signal estimates. In the context of target series obtained from repeated surveys, a design-based version of SKF was also proposed under a working model. This should have important practical applications because it ensures design-consistent benchmarked signal estimates even though the working model assumptions may not be correct. Finally, we remark that the basic idea underlying SKF is quite general and can be used in problems of estimating linear combinations of parameters under linear constraints. In particular, the problems of calendarization, interpolation, and forecasting can be unified with the problem of benchmarking, see also Dagum, Cholette, and Chen (1996).

REFERENCES

Bell, W. (1984). Signal extraction for nonstationary time series. *The Annals of Statistics*, 12, 646-664.

Chen, Z.-G., Cholette, P.A., and Dagum, E.B. (1995). A non-parametric method for benchmarking survey data via signal extraction. Technical Report, Business Survey Methods Division, Statistics Canada, Ottawa.

Cholette, P.A., and Dagum, E.B. (1994). Benchmarking time series with auto-correlated sampling errors. *International Statistical Review*, 62, 365-377.

Dagum, E.B., Cholette, P.A., and Chen, Z.-G. (1996). A unified view of signal extraction, benchmarking, and interpolation of time series. Technical Report, Business Survey Methods Division, Statistics Canada, Ottawa.

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

Durbin, J., and Quenneville, B. (1997). Benchmarking by state space models. *International Statistical Review*, (to appear).

Harvey, A.C. (1989). Forecasting, structural time series models, and the Kalman filter. Cambridge Univ. Press, Cambridge.

Hillmer, C.H., and Trabelsi, A. (1987). Benchmarking of economic time series. *Jour. Amer. Stat. Assoc.*, 82, 1064-1071.

Mian, I.U.H., and Laniel, N. (1993). Maximum likelihood estimation of constant multiplicative bias benchmarking model with application. *Survey Methodology*, 19, 165-172.

Odell, P.L., and Lewis, T.O. (1971). Best Linear Recursive Estimation. *Jour. Amer. Stat.*

Assoc., 66, 893-896.

Pfeffermann, D., and Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology*, 16, 217-237.

Rao, C.R. (1968). A note on a previous lemma in the theory of least squares and some further results. *Sankhyā A* 30, 259-266.

Singh, A.C. (1995a). Predicting functions for generalization of BLUP to mixed nonlinear models. *ASA Proc. Biom. Sec.* (to appear).

Singh, A.C. (1995b). Discussion of Papers by Durbin/Quenneville and Bell. *SSC Proc. Surv. Meth. Sec.* (pp 29-31).

Stone, J.R.N., Champernowne, D.G., and Meade, J.E. (1942). The precision of national income estimates. *Review of Economic Studies*, 9, 111-125.

Table 1(a). Benchmarked y-series with nonrandom signals (For nonrandom annual and monthly benchmarks)

t	y_{1t}	$\hat{\eta}_{1t}^{(1)}$	$\hat{\eta}_{1t}^{(2)}$	$\hat{\eta}_{1t}^{(3)}$	$\hat{\eta}_{1t}^{(4)}$	$\hat{\eta}_{1t}^{(5)}$	y_{2t}	$\hat{\eta}_{2t}^{(1)}$	$\hat{\eta}_{2t}^{(2)}$	$\hat{\eta}_{2t}^{(3)}$	$\hat{\eta}_{2t}^{(4)}$	$\hat{\eta}_{2t}^{(5)}$	z_t
1	402.37	444.76	444.76	469.22	485.04	485.68	767.51	817.80	817.80	923.19	920.48	935.42	1392.41
2	423.96	466.35	466.35	435.34	450.49	451.10	1010.97	1061.27	1061.27	939.77	937.18	951.49	1375.11
3	363.51	405.90	405.90	402.47	416.02	416.57	927.79	978.08	978.08	1032.89	1030.57	1043.37	1435.37
4	438.46	480.85	480.85	445.18	456.94	457.41	1135.68	1185.97	1185.97	1004.71	1002.70	1013.80	1449.89
5	381.17	423.56	423.56	398.26	406.52	406.86	1086.68	1136.98	1136.98	1033.90	1032.49	1040.29	1432.16
6	352.16	394.55	394.55	389.80	394.69	394.89	984.74	1035.04	1035.04	1119.69	1118.85	1123.47	1509.48
7	306.70	349.09	349.09	344.71	346.17	346.23	1017.41	1067.71	1067.71	1155.57	1155.33	1156.70	1500.28
8	467.40	509.79	509.79	485.18	483.51	483.44	1236.43	1286.73	1286.73	1124.47	1124.76	1123.18	1609.65
9	242.93	285.32	285.32	317.12	307.47	307.08	975.76	1026.05	1026.05	1307.54	1309.19	1300.07	1624.65
10	437.55	479.94	479.94	490.20	477.96	477.47	1164.18	1214.48	1214.48	1105.21	1107.30	1095.74	1595.41
11	320.14	362.53	362.53	400.48	379.49	378.64	1045.11	1095.40	1095.40	1206.49	1210.08	1190.26	1606.97
12	309.82	352.21	352.21	376.89	350.55	349.49	1208.99	1259.28	1259.28	1211.35	1215.86	1190.98	1588.24
Annual Sum	4446.17	4954.85	4954.85	4954.85	4954.85	4954.85	12561.25	13164.79	13164.79	13164.78	13164.79	13164.77	18119.62
(x_{js})		(x_{11})	(x_{11})	(x_{11})	(x_{11})	(x_{11})		(x_{21})	(x_{21})	(x_{21})	(x_{21})	(x_{21})	18119.62
13	397.97	440.37	440.37	465.04	432.78	431.47	1150.28	1200.58	1200.58	1152.64	1174.52	1053.83	1485.31
14	438.37	480.76	480.76	505.43	464.21	496.85	769.90	820.20	820.20	772.26	808.56	930.54	1427.40
15	343.75	386.14	386.14	410.81	359.71	329.62	1170.33	1220.63	1220.63	1172.69	1214.86	1088.33	1417.95
16	281.87	324.26	324.26	348.93	292.36	284.66	974.74	1025.03	1025.03	977.10	1031.47	1132.31	1416.96
17	394.79	437.18	437.18	461.85	402.02	382.90	1203.41	1253.70	1253.70	1205.77	1267.67	1175.26	1558.16
18	307.21	349.60	349.60	374.27	308.83	301.57	1165.69	1215.98	1215.98	1168.05	1239.98	1218.51	1520.08
19	326.90	369.29	369.29	393.96	325.60	329.40	1017.88	1068.17	1068.17	1020.24	1100.25	1215.10	1544.50
20	262.45	304.84	304.84	329.51	258.41	262.57	1094.94	1145.24	1145.24	1097.30	1182.44	1274.19	1536.77
21	454.04	496.43	496.43	521.10	448.58	445.41	1088.90	1139.20	1139.20	1091.26	1181.15	1104.21	1549.61
22	435.35	477.74	477.74	502.41	426.71	448.39	867.53	917.82	917.82	869.89	963.30	1028.48	1476.87
23	489.41	531.80	531.80	556.47	478.82	479.67	945.65	995.95	995.95	948.01	1042.92	1007.24	1486.91
24	392.46	434.85	434.85	459.52	380.63	386.13	1064.15	1114.45	1114.45	1066.51	1162.30	1141.43	1527.56
Annual Sum	4524.57	5033.26	5033.26	5329.33	4578.66	4578.66	12513.40	13116.95	13116.95	12541.72	13369.42	13369.43	17948.08
(x_{js})					(x_{12})	(x_{12})					(x_{22})	(x_{22})	17948.08

Table 1(b). Benchmarked y-series with serially dependent signals
(For nonrandom annual and monthly benchmarks)

t	y_{1t}	$\hat{\eta}_{1t}^{(1)}$	$\hat{\eta}_{1t}^{(2)}$	$\hat{\eta}_{1t}^{(3)}$	$\hat{\eta}_{1t}^{(4)}$	$\hat{\eta}_{1t}^{(5)}$	y_{2t}	$\hat{\eta}_{2t}^{(1)}$	$\hat{\eta}_{2t}^{(2)}$	$\hat{\eta}_{2t}^{(3)}$	$\hat{\eta}_{2t}^{(4)}$	$\hat{\eta}_{2t}^{(5)}$	z_t
1	402.37	402.37	461.80	477.28	479.09	480.17	767.51	767.51	825.96	915.15	913.33	912.26	1392.42
2	423.96	386.05	424.90	410.69	411.29	411.64	1010.97	995.10	1047.82	964.42	963.82	963.47	1375.11
3	363.51	371.14	414.99	409.31	409.92	410.29	927.79	954.15	1017.73	1026.06	1025.45	1025.08	1435.37
4	438.46	394.68	441.72	419.91	420.52	420.88	1135.68	1039.53	1109.23	1029.98	1029.37	1029.01	1449.89
5	381.17	371.47	420.34	399.17	399.75	400.10	1086.68	1040.90	1113.94	1032.99	1032.41	1032.06	1432.17
6	352.16	361.48	411.15	410.54	411.06	411.38	984.74	1016.85	1091.35	1098.94	1098.42	1098.10	1509.49
7	306.70	342.32	391.93	388.73	389.15	389.40	1017.41	1051.81	1126.26	1111.55	1111.13	1110.88	1500.28
8	467.40	393.04	441.76	440.50	440.74	440.87	1236.43	1111.87	1184.86	1169.15	1168.92	1168.78	1609.66
9	242.93	318.99	365.89	390.58	390.49	390.44	975.76	1066.37	1136.40	1234.08	1234.16	1234.21	1624.66
10	437.55	383.69	427.61	431.65	431.02	430.64	1164.18	1100.70	1166.06	1163.76	1164.39	1164.77	1595.41
11	320.14	344.78	384.20	401.71	400.16	399.23	1045.11	1085.20	1143.82	1205.26	1206.81	1207.74	1606.97
12	309.82	335.70	368.51	374.79	371.66	369.80	1208.99	1152.07	1201.42	1213.45	1216.58	1218.44	1588.23
Annual Sum	4446.17	4405.71	4954.80	4954.83	4954.83	4954.83	12561.25	12382.06	13164.85	13164.79	13164.79	13164.80	18119.66
(x_{js})			(x_{11})	(x_{11})	(x_{11})	(x_{11})			(x_{21})	(x_{21})	(x_{21})	(x_{21})	18119.64
13	397.97	372.76	395.94	400.47	398.97	382.71	1150.28	1089.48	1126.48	1135.25	1162.19	1102.60	1485.30
14	438.37	401.46	417.92	421.20	421.60	420.65	769.90	913.28	940.81	947.20	989.75	1006.75	1427.40
15	343.75	347.56	359.30	361.67	363.83	337.32	1170.33	1101.77	1122.14	1126.78	1179.61	1080.63	1417.94
16	281.87	329.82	338.21	339.93	343.53	328.36	974.74	1057.88	1072.88	1076.26	1135.79	1088.60	1416.96
17	394.79	362.79	368.81	370.05	374.73	366.43	1203.41	1135.66	1146.67	1149.12	1212.88	1191.73	1558.16
18	307.21	332.97	337.30	338.20	343.60	333.63	1165.69	1147.57	1155.63	1157.41	1223.59	1186.45	1520.08
19	326.90	348.14	351.25	351.91	357.69	362.99	1017.88	1082.53	1088.42	1089.71	1156.84	1181.51	1544.50
20	262.45	326.85	329.10	329.57	335.44	339.89	1094.94	1113.31	1117.61	1118.55	1185.29	1196.88	1536.76
21	454.04	404.57	406.20	406.54	412.22	421.13	1088.90	1035.75	1038.88	1039.56	1104.47	1128.48	1549.62
22	435.35	409.42	410.59	410.84	416.14	434.37	867.53	918.42	920.70	921.19	982.49	1042.50	1476.87
23	489.41	423.64	424.49	424.67	429.47	448.00	945.65	925.18	926.84	927.20	982.42	1038.91	1486.90
24	392.46	376.22	376.83	376.96	381.38	403.15	1064.15	1007.10	1008.31	1008.56	1054.10	1124.41	1527.56
Annual Sum	4524.57	4436.20	4515.95	4532.00	4578.60	4578.61	12513.40	12527.93	12665.37	12696.79	13369.42	13369.45	17948.05
(x_{js})					(x_{12})	(x_{12})					(x_{22})	(x_{22})	17948.03

Table 2(a). **Benchmarked y-series with nonrandom signals**
(For nonrandom annual and dependent but nonbinding monthly benchmarks)

t	y_{1t}	$\hat{\eta}_{1t}^{(1)}$	$\hat{\eta}_{1t}^{(2)}$	$\hat{\eta}_{1t}^{(3)}$	$\hat{\eta}_{1t}^{(4)}$	$\hat{\eta}_{1t}^{(5)}$	y_{2t}	$\hat{\eta}_{2t}^{(1)}$	$\hat{\eta}_{2t}^{(2)}$	$\hat{\eta}_{2t}^{(3)}$	$\hat{\eta}_{2t}^{(4)}$	$\hat{\eta}_{2t}^{(5)}$	z_t
1	402.37	444.76	444.76	470.34	486.54	486.91	767.51	817.80	817.80	920.52	918.91	934.57	1389.78
2	423.96	466.35	466.35	435.28	450.80	451.15	1010.97	1061.27	1061.27	932.31	930.77	945.78	1365.33
3	363.51	405.90	405.90	400.61	414.48	414.80	927.79	978.08	978.08	1013.84	1012.46	1025.89	1412.68
4	438.46	480.85	480.85	444.21	456.25	456.53	1135.68	1185.97	1185.97	992.26	991.06	1002.72	1433.69
5	381.17	423.56	423.56	396.79	405.25	405.45	1086.68	1136.98	1136.98	1017.65	1016.81	1025.02	1412.06
6	352.16	394.55	394.55	389.95	394.96	395.08	984.74	1035.04	1035.04	1118.27	1117.77	1122.65	1505.98
7	306.70	349.09	349.09	344.85	346.34	346.38	1017.41	1067.71	1067.71	1154.06	1153.91	1155.38	1496.22
8	467.40	509.79	509.79	486.28	484.57	484.53	1236.43	1286.73	1286.73	1134.89	1135.06	1133.42	1618.09
9	242.93	285.32	285.32	317.06	307.18	306.95	975.76	1026.05	1026.05	1308.84	1309.82	1300.28	1624.32
10	437.55	479.94	479.94	490.84	478.31	478.02	1164.18	1214.48	1214.48	1119.24	1120.49	1108.35	1605.95
11	320.14	362.53	362.53	400.23	378.74	378.25	1045.11	1095.40	1095.40	1213.45	1215.59	1194.76	1612.50
12	309.82	352.22	352.22	378.41	351.43	350.81	1208.99	1259.28	1259.28	1239.43	1242.12	1215.95	1614.19
Annual Sum	4446.17	4954.85	4954.85	4954.85	4954.85	4954.85	12561.25	13164.79	13164.79	13164.76	13164.77	13164.77	18090.79
(x_{js})			(x_{11})	(x_{11})	(x_{11})	(x_{11})			(x_{21})	(x_{21})	(x_{21})	(x_{21})	18119.64
13	397.97	440.37	440.37	466.56	433.52	432.76	1150.28	1200.58	1200.58	1180.73	1193.71	1067.37	1505.84
14	438.37	480.76	480.76	506.95	464.73	498.01	769.90	820.20	820.20	800.35	821.87	934.00	1440.59
15	343.75	386.14	386.14	412.34	359.99	327.38	1170.33	1220.63	1220.63	1200.78	1225.78	1076.42	1413.10
16	281.87	324.26	324.26	350.45	292.50	283.12	974.74	1025.03	1025.03	1005.19	1037.42	1121.96	1416.67
17	394.79	437.18	437.18	463.38	402.08	382.16	1203.41	1253.70	1253.70	1233.86	1270.55	1170.98	1564.95
18	307.21	349.60	349.60	375.79	308.76	300.98	1165.69	1215.98	1215.98	1196.13	1238.78	1209.64	1523.49
19	326.90	369.29	369.29	395.48	325.46	330.13	1017.88	1068.17	1068.17	1048.32	1095.75	1222.65	1566.02
20	262.45	304.84	304.84	331.04	258.20	263.58	1094.94	1145.24	1145.24	1125.39	1175.86	1281.59	1557.89
21	454.04	496.43	496.43	522.63	448.34	446.61	1088.90	1139.20	1139.20	1119.35	1172.63	1112.54	1571.21
22	435.35	477.74	477.74	503.93	426.39	448.99	867.53	917.82	917.82	897.97	953.34	1030.74	1491.99
23	489.41	531.80	531.80	557.99	478.45	479.27	945.65	995.95	995.95	976.10	1032.35	1004.10	1492.65
24	392.46	434.85	434.85	461.04	380.24	385.66	1064.15	1114.45	1114.45	1094.60	1151.38	1137.45	1531.14
Annual Sum	4524.57	5033.26	5033.26	5347.58	4578.66	4578.66	12513.40	13116.95	13116.95	12878.77	13369.42	13369.44	18075.54
(x_{js})					(x_{12})	(x_{12})					(x_{22})	(x_{22})	17948.03

Table 2(b). Benchmarked y-series with serially dependent signals (For nonrandom annual and serially dependent but nonbinding monthly benchmarks)

t	y_{1t}	$\hat{\eta}_{1t}^{(1)}$	$\hat{\eta}_{1t}^{(2)}$	$\hat{\eta}_{1t}^{(3)}$	$\hat{\eta}_{1t}^{(4)}$	$\hat{\eta}_{1t}^{(5)}$	y_{2t}	$\hat{\eta}_{2t}^{(1)}$	$\hat{\eta}_{2t}^{(2)}$	$\hat{\eta}_{2t}^{(3)}$	$\hat{\eta}_{2t}^{(4)}$	$\hat{\eta}_{2t}^{(5)}$	z_t
1	402.37	402.37	461.79	476.32	478.04	479.50	767.51	767.51	825.96	910.13	907.39	908.94	1389.78
2	423.96	386.05	424.90	409.10	409.39	410.70	1010.97	995.10	1047.82	968.37	966.75	969.43	1365.33
3	363.51	371.14	414.99	404.59	404.85	406.32	927.79	954.15	1017.73	1003.15	1001.38	1004.45	1412.68
4	438.46	394.68	441.72	419.39	419.63	421.14	1135.68	1039.53	1109.23	1029.74	1027.95	1031.11	1433.69
5	381.17	371.47	420.34	398.84	399.08	400.50	1086.68	1040.90	1113.94	1032.63	1030.94	1033.91	1412.06
6	352.16	361.48	411.14	410.34	410.58	411.78	984.74	1016.85	1091.35	1098.29	1096.82	1099.32	1505.98
7	306.70	342.32	391.93	389.41	389.64	390.46	1017.41	1051.81	1126.26	1114.08	1112.99	1114.70	1496.22
8	467.40	393.04	441.76	443.13	443.31	443.60	1236.43	1111.87	1184.86	1180.43	1179.93	1180.52	1618.09
9	242.93	318.99	365.89	388.31	388.35	387.89	975.76	1066.37	1136.40	1222.12	1222.52	1221.58	1624.32
10	437.55	383.69	427.61	434.19	433.91	432.43	1164.18	1100.70	1166.06	1173.59	1175.32	1172.36	1605.95
11	320.14	344.78	384.20	401.67	400.72	397.87	1045.11	1085.20	1143.82	1201.62	1205.36	1199.82	1612.50
12	309.82	335.70	368.50	379.48	377.24	372.55	1208.99	1152.07	1201.42	1230.69	1237.50	1228.71	1614.19
Annual Sum	4446.17	4405.71	4954.80	4954.80	4954.80	4954.80	12561.25	12382.06	13164.84	13164.84	13164.84	13164.84	18090.79
(x_{js})			(x_{11})	(x_{11})	(x_{11})	(x_{11})			(x_{21})	(x_{21})	(x_{21})	(x_{21})	18119.64
13	397.97	372.76	395.93	404.17	402.83	386.36	1150.28	1089.48	1126.48	1147.16	1174.94	1116.39	1505.84
14	438.37	401.46	417.92	424.05	424.03	421.09	769.90	913.28	940.81	955.50	996.95	1005.89	1440.59
15	343.75	347.56	359.29	363.83	365.13	338.84	1170.33	1101.77	1122.14	1132.61	1182.99	1085.36	1413.10
16	281.87	329.82	338.21	341.55	343.96	328.98	974.74	1057.88	1072.88	1080.37	1136.50	1089.51	1416.67
17	394.79	362.79	368.81	371.26	374.53	367.20	1203.41	1135.66	1146.67	1152.04	1211.83	1193.65	1564.95
18	307.21	332.97	337.29	339.09	342.92	334.23	1165.69	1147.57	1155.63	1159.49	1221.32	1187.76	1523.49
19	326.90	348.14	351.25	352.56	356.69	364.14	1017.88	1082.53	1088.42	1091.19	1153.76	1185.72	1566.02
20	262.45	326.85	329.09	330.05	334.23	341.08	1094.94	1113.31	1117.61	1119.61	1181.72	1202.04	1557.89
21	454.04	404.57	406.19	406.89	410.92	421.59	1088.90	1035.75	1038.88	1040.32	1100.68	1131.76	1571.21
22	435.35	409.42	410.59	411.10	414.82	432.26	867.53	918.42	920.70	921.74	978.71	1035.88	1491.99
23	489.41	423.64	424.49	424.86	428.23	444.25	945.65	925.18	926.84	927.60	978.89	1026.48	1492.65
24	392.46	376.22	376.81	377.09	380.30	398.57	1064.15	1007.10	1008.31	1008.85	1051.10	1109.01	1531.14
Annual Sum	4524.57	4436.20	4515.94	4546.54	4578.60	4578.60	12513.40	12527.93	12665.37	12736.48	13369.44	13369.45	18075.54
(x_{js})					(x_{12})	(x_{12})					(x_{22})	(x_{22})	17948.03

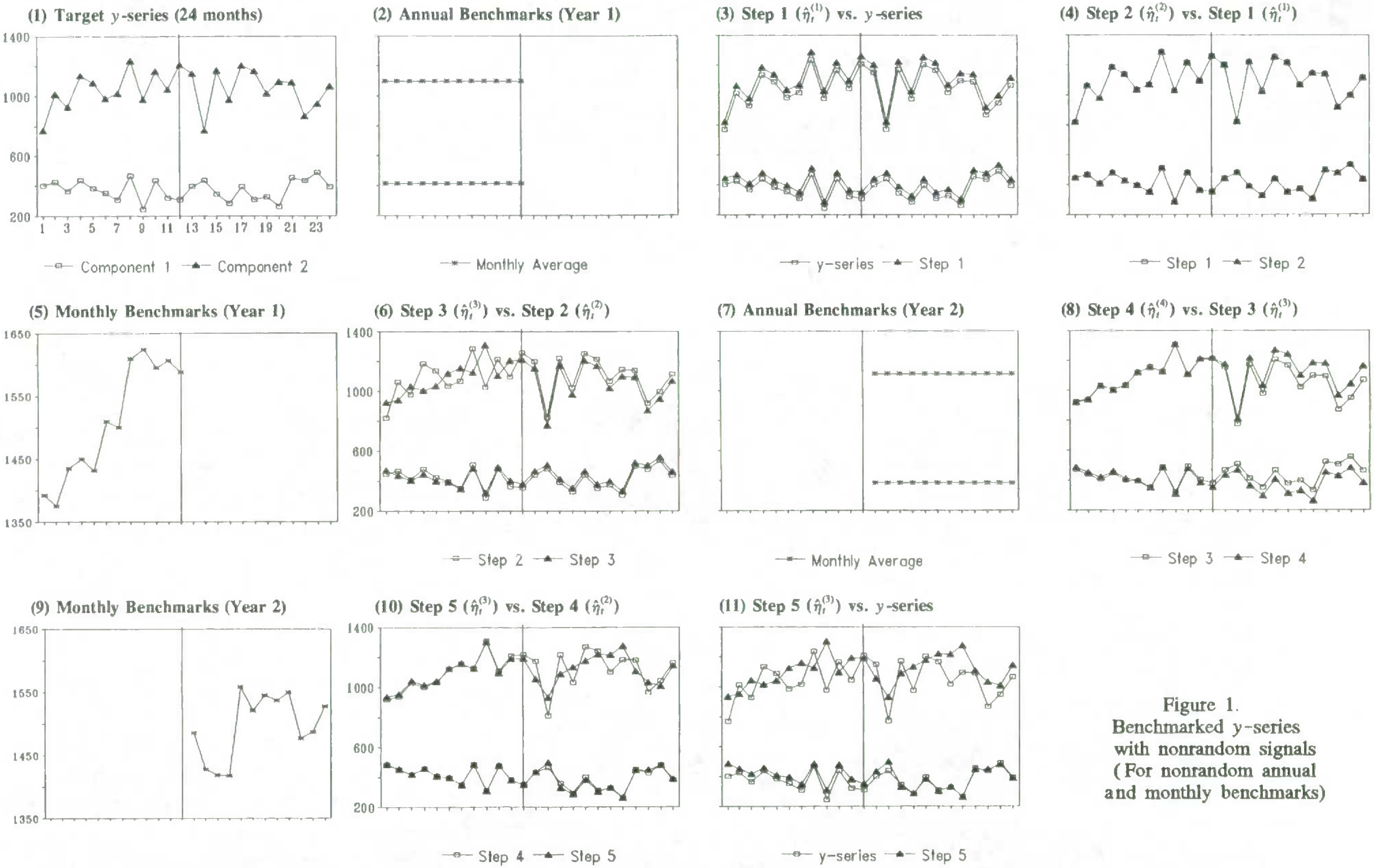

Figure 1.
Benchmarked y-series with nonrandom signals (For nonrandom annual and monthly benchmarks)

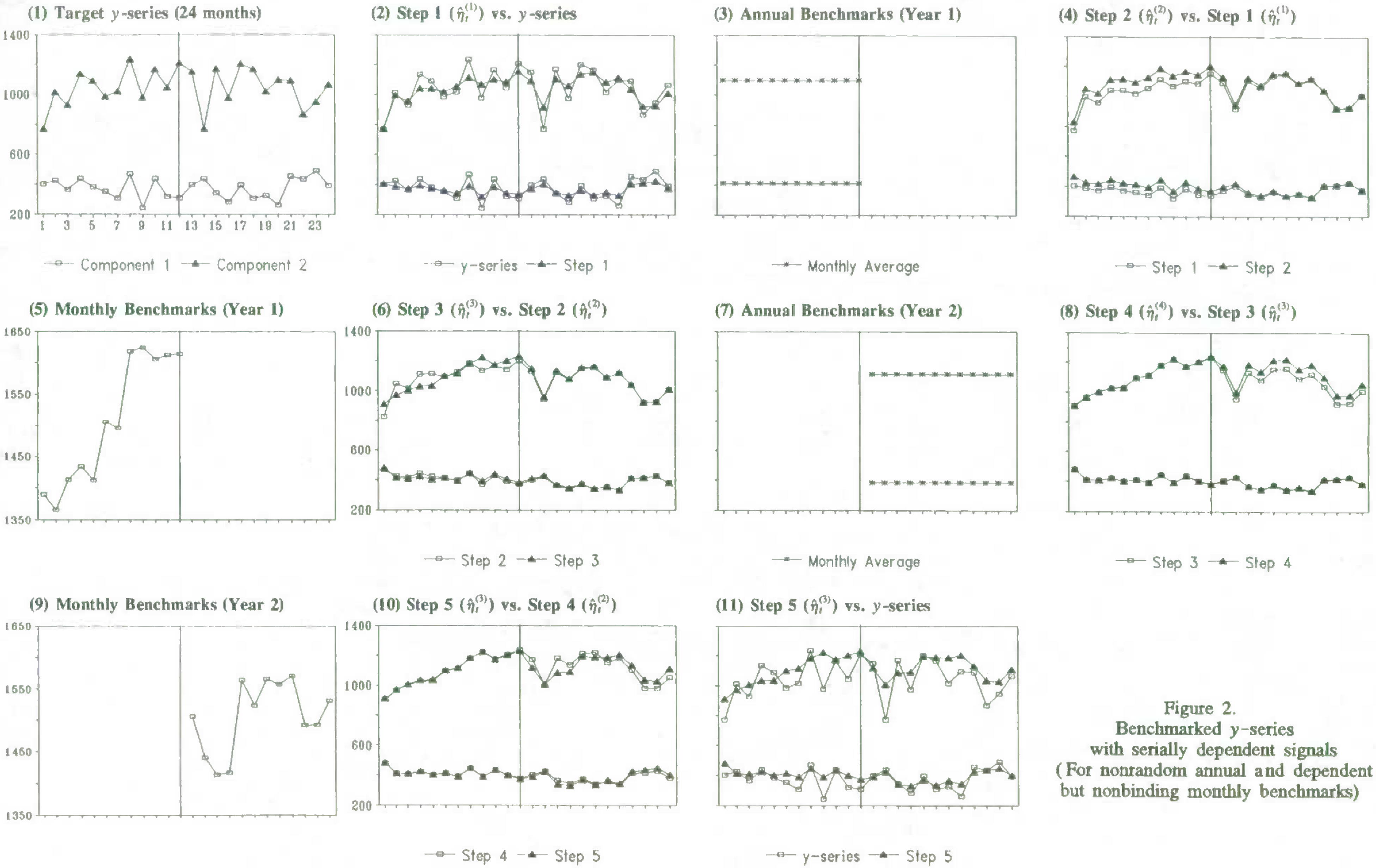

Figure 2.
Benchmarked y-series
with serially dependent signals
(For nonrandom annual and dependent
but nonbinding monthly benchmarks)