Journal of Forecasting
J. Forecast. **19**, 103–122 (2000)

Linear Combination of Restrictions and Forecasts in Time Series Analysis

VICTOR M. GUERRERO[1] AND DANIEL PEÑA[2]*
[1] *Instituto Tecnológico Autónomo de México (ITAM), México*
[2] *Universidad Carlos III de Madrid, Spain*

ABSTRACT

An important tool in time series analysis is that of combining information in an optimal way. Here we establish a basic combining rule of linear predictors and show that such problems as forecast updating, missing value estimation, restricted forecasting with binding constraints, analysis of outliers and temporal disaggregation can be viewed as problems of optimal linear combination of restrictions and forecasts. A compatibility test statistic is also provided as a companion tool to check that the linear restrictions are compatible with the forecasts generated from the historical data. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS Compatibility testing; disaggregation; missing data; outliers; restricted forecasts

INTRODUCTION

Combining information is so common in the practice of statistics that the practising statistician often does not realize that he or she is applying it. Hedges and Olkin (1985) present many statistical problems that can be analysed from this point of view. Draper *et al*. (1992) provide a thorough review of this field with many examples and ideas for future research. Similarly, Peña (1997) considers combining information with emphasis on understanding the structure and properties of the estimators involved in the combination. In fact, the idea of combining observations can be traced back to Gauss (see Young, 1984).

This paper presents a basic (least squares) rule that has been frequently used in time series for combining information. We consider here that some information, additional to the time series data, is available in the form of linear restrictions that have to be fulfilled exactly by an optimal predictor. Our basic concern is to obtain (conditionally) unbiased Minimum Mean Square Error Linear Predictors (MMSELP) of random vectors. Hence no distributional assumption will be

* Correspondence to: Professor Daniel Peña, Departmento de Estadística y Econometría, Universidad Carlos III de Madrid, Getafe, Madrid 28903, Spain. E-mail: dpena@est-econ.uc3m.es

Contract grant sponsor: Catedra BBV de Calidad.
Contract grant sponsor: DGES.
Contract grant number: PB96-0111.

required for obtaining the optimal predictors, although when normality is a reasonable assumption, the linear qualification can be dropped from MMSELP.

Each of the two sources of information is assumed to provide a linear and (conditionally) unbiased predictor. The unbiasedness assumption may be considered in some instances as unduly restrictive (see Palm and Zellner, 1992 or Min and Zellner, 1993). For our purposes and in the problems here considered, we deem such an assumption general enough and unrestrictive, since debiasing can be carried out before combining. With respect to the use of only two sources of information, we note that this is only to ease the exposition, since the ideas can be extended straightforwardly to several sources.

We assume that the models involved as well as their corresponding parameters are known, so that model building and parameter estimation are of no importance for us. Neither are we concerned with such an issue as efficient computer implementation of theoretical solutions provided by the combining rule. In fact, we are mainly interested in emphasizing the central role played by the basic rule, as a unifying tool of several apparently different approaches that have been employed to provide techniques in linear time series analysis. Furthermore, throughout this paper we consider specifically the families of autoregressive integrated moving average (ARIMA) models to represent the behaviour of a univariate time series. Nevertheless, the results hold true for any linear time series model. It should be stressed that we do not claim originality in the solution of the problems considered here, since they have been solved by many different authors, including ourselves in previous work. In fact, we do not provide real data examples, since they can be found in the original papers dealing with the particular problems encompassed by our approach. What we think is new is to show that such problems as missing data, restricted forecasting with binding constraints, influential outliers and temporal disaggregation, among others, can be seen as particular cases of a simple structure and therefore they share the same basic solution. This result can be useful to transfer optimal solutions found for any of these problems to other fields. For instance, issues of robustness, non-normality or non-linearity could be seen under the same structure.

In the following section we establish the notation and the basic combining rule which will be used extensively in subsequent sections. A test statistic for validating an implicit compatibility assumption between sources of information is also provided there. After that, we apply and interpret the basic rule within the context of forecast updating and missing data estimation. Then we concentrate on the problem of restricted forecasting, with some variants that respond to different states of knowledge about the future. Afterwards we touch upon the problems of influential and reallocation outliers in time series. Several simulated numerical examples are provided to illustrate the use of the rule in practical applications. The final section concludes with some comments and points to the need for some other combining rules.

BASIC COMBINING RULE

Here we present an optimal combining rule that can be employed when two basic sources of information are available: (1) a statistical model based on an observed data set X, that produces the unrestricted MMSELP $\hat{\mathbf{Z}}_p$, of the random vector $\mathbf{Z}$, and (2) some extra-model information $\mathbf{Y} = C\mathbf{Z}$ given in the form of linear restrictions imposed on $\mathbf{Z}$. As indicated in the Introduction, we shall assume that the model is known, as well as its parameters. We now establish the rule and illustrate its use in different situations afterwards.

BASIC COMBINING RULE (BCR). Let us suppose that the two observed vectors $\hat{\mathbf{Z}}_p$, of dimension $h \times 1$, and $\mathbf{Y}$, of dimension $m \times 1$, are related to an unobserved random variable $\mathbf{Z}$ by

$$\hat{\mathbf{Z}}_p = E(\mathbf{Z} \mid X) \tag{1}$$

where X is a well-defined information set, and

$$\mathbf{Y} = C\mathbf{Z} \tag{2}$$

where C is a known $m \times h$ matrix of rank $m < h$. Let $\mathbf{e} = \mathbf{Z} - \hat{\mathbf{Z}}_p$ be the forecast error and assume that $E(\mathbf{e} \mid X) = \mathbf{0}$, $E(\mathbf{Z}\mathbf{e}' \mid X) = 0$ and $\mathrm{Cov}(\mathbf{e} \mid X) = \Sigma_e$ with Σ_e a known positive definite matrix. Then the MMSELP of $\mathbf{Z}$, based on $\hat{\mathbf{Z}}_p$ and $\mathbf{Y}$, is given by

$$\hat{\mathbf{Z}} = \hat{\mathbf{Z}}_p + A(\mathbf{Y} - C\hat{\mathbf{Z}}_p) \tag{3}$$

where

$$A = \Sigma_e C'(C\Sigma_e C')^{-1} \tag{4}$$

Furthermore, the MSE matrix of errors for this adjusted forecast, $\Gamma = \mathrm{Cov}(\hat{\mathbf{Z}} - \mathbf{Z})$, is given by

$$\Gamma = (I_h - AC)\Sigma_e \tag{5}$$

where I_h is the h-dimensional identity matrix.
Proof Any linear predictor of $\mathbf{Z}$ based on $\mathbf{Y}$ and $\hat{\mathbf{Z}}_p$ must be the form

$$\tilde{\mathbf{Z}} = A_1\mathbf{Y} + A_2\hat{\mathbf{Z}}_p = (A_1C + A_2)\hat{\mathbf{Z}}_p + A_1C\mathbf{e}$$

with A_1 and A_2 some constant matrices. The forecast error is $\tilde{\mathbf{Z}} - \mathbf{Z} = (A_1C + A_2 - I_h)\hat{\mathbf{Z}}_p - (I_h - A_1C)\mathbf{e}$, and we shall require that $\tilde{\mathbf{Z}}$ satisfy Gauss's error consistency (see Sprott, 1978) which leads us to select $\tilde{\mathbf{Z}}$ as the predictor that coincides with $\mathbf{Z}$ when $\mathbf{e} = \mathbf{0}$. In such a case

$$A_1C + A_2 = I_h$$

which implies conditional unbiasedness since $E(\tilde{\mathbf{Z}} - \mathbf{Z} \mid X) = \mathbf{0}$. Then $\tilde{\mathbf{Z}} - \mathbf{Z} = -(I_h - A_1C)\mathbf{e}$ and

$$\mathrm{Cov}(\tilde{\mathbf{Z}} - \mathbf{Z}) = (I_h - A_1C)\Sigma_e(I_h - A_1C)'$$

Therefore, by taking the derivative of this equation with respect to A_1 and equating it to zero we obtain $A_1 = A$. Alternatively, let us consider $A_1 = A + \Delta$ with Δ an arbitrary constant matrix, then it follows that

$$\mathrm{Cov}(\tilde{\mathbf{Z}} - \mathbf{Z}) = \Sigma_e - \Sigma_e C'(C\Sigma_e C')^{-1}C\Sigma_e + \Delta C\Sigma_e C'\Delta' \tag{6}$$

Hence, the MSE matrix $\mathrm{Cov}(\tilde{\mathbf{Z}} - \mathbf{Z})$ of any linear and (conditionally) unbiased predictor $\tilde{\mathbf{Z}}$ exceeds Γ by a positive semidefinite matrix ■

Comments

(1) As an example of the BCR, let us consider the original Kalman state-space model (see Young, 1984) given by the state equation

$$\mathbf{Z}_t = \Phi \mathbf{Z}_{t-1} + \mathbf{a}_t$$

and a measurement equation without noise

$$\mathbf{Y}_t = C\mathbf{Z}_t$$

At time $t-1$, before observing $\mathbf{Y}_t$ we have an estimate of the state $\mathbf{Z}_t$ given by $\hat{\mathbf{Z}}_{t|t-1} = \Phi\hat{\mathbf{Z}}_{t-1}$, where $\hat{\mathbf{Z}}_{t-1}$ is the estimate of $\mathbf{Z}_{t-1}$ including the information up to $\mathbf{Y}_{t-1}$. Let us call $P_{t\,|\,t-1}$ to the covariance matrix of this estimate. Once $\mathbf{Y}_t$ is observed, the previous estimate is revised so that the restriction is fulfilled. Then, by using equation (3) with $\Sigma_e = P_{t|t-1}$, $\hat{\mathbf{Z}}_p = \Phi\hat{\mathbf{Z}}_{t-1}$, $\mathbf{Y} = \mathbf{Y}_t$ and $\hat{\mathbf{Z}} = \hat{\mathbf{Z}}_{t\,|\,t}$, we have that

$$\hat{\mathbf{Z}}_{t|t} = \hat{\mathbf{Z}}_{t|t-1} + A_t(\mathbf{Y}_t - C\hat{\mathbf{Z}}_{t|t-1})$$

where $A_t = P_{t|t-1}C'(CP_{t|t-1}C')^{-1}$. This is the standard state updating equation of the Kalman Filter and A_t is the Kalman Gain. We conclude that the state-space model without measurement error is a particular case of the formulation presented in the paper. The formulation of the BCR does not assume any particular structure for the form of the forecast $\hat{\mathbf{Z}}_p$, whereas in the state-space model the Markovian structure imposed by the state equation leads to efficient recursive computations.

(2) The estimate of $\mathbf{Z}$ given by equation (3) can be interpreted as a linear combination of the linear and (conditionally) unbiased predictors provided by the two different sources of information. To show this, note that equation (2) leads to the linear predictor $\hat{\mathbf{Z}}_1 = P\mathbf{Y}$ with P a matrix satisfying the condition $CP = I_m$, so that $C\hat{\mathbf{Z}}_1$ coincides with $\mathbf{Y}$. Similarly, $\hat{\mathbf{Z}}_p$ is a linear and (conditionally) unbiased predictor of $\mathbf{Z}$. Then, we should combine linearly $\hat{\mathbf{Z}}_1$ and $\hat{\mathbf{Z}}_p$ in such a way that the MSE matrix of the resulting predictor is minimized, and we go back to equation (3).

(3) We can allow Σ_e to be singular, provided that $C\Sigma_e C'$ is non-singular. This is a useful extension, since it enables iterative application of the BCR, e.g. given a known constraint $\mathbf{Y}_1 = C_1\mathbf{Z}$, one could consider a further constraint $\mathbf{Y}_2 = C_2\mathbf{Z}$ provided that (C_1, C_2) is a full rank matrix.

(4) Γ is always singular because the estimator $\hat{\mathbf{Z}}$ satisfies the linear restriction (2), thus $C\Gamma = 0$ and Γ has rank $h - m$.

(5) The formulation considered can be generalized by including an additive error term in equation (2). In that case, the restrictions would be considered as subjected to uncertainty and the problem lies in combining two forecasts. Several time series problems, such as benchmarking (see, for instance, Hillmer and Trabelsi, 1987, or Cholette and Dagum, 1994) can be covered by this umbrella. However, in this paper we concentrate on the specific interesting case in which the restriction is deemed certain and must be fulfilled exactly.

(6) An alternative specification would consider a prior distribution for $\mathbf{Y}$. In that case, a complete Bayesian analysis is advisable. Some work dealing with that approach has already appeared in the literature. We refer the reader to de Alba (1988, 1993) for more information on this topic.
(7) The closed expressions of the BCR provide a fairly simple and straightforward theoretical solution to a very general problem. However, when the dimension h is high the solution may not be computationally efficient and it is more convenient to formulate the problem in a recursive way to compute the solution. In fact, in several of the problems we will analyse in the following sections the general solution presented can be computed by a recursive method such as Kalman filtering.

The BCR allows us to combine $\hat{\mathbf{Z}}_p$ and $\mathbf{Y}$ in an optimal manner, but it does not necessarily follow that $\hat{\mathbf{Z}}_p$ and $\mathbf{Y}$ should always be combined. In particular, it will not be sensible to combine them when they contradict each other. Then, it makes sense to test if the restriction (2) is compatible with the data set that produced $\hat{\mathbf{Z}}_p$. To that end, a compatibility test (CT) derived on the assumption of normality for $\mathbf{e}$ can be employed. That is, let us consider as null hypothesis H_0: $\mathbf{Y} - C\mathbf{Z} = \mathbf{0}$. On this hypothesis, $\mathbf{Y} - C\hat{\mathbf{Z}}_p$ is normally distributed with mean vector zero and covariance matrix $C\Sigma_e C'$. Therefore, a statistic for testing the compatibility between $\mathbf{Y}$ and $\hat{\mathbf{Z}}_p$ is given by

$$K = (\mathbf{Y} - C\hat{\mathbf{Z}}_p)'(C\Sigma_e C')^{-1}(\mathbf{Y} - C\hat{\mathbf{Z}}_p) \tag{7}$$

which will be distributed as a χ^2 with m degrees of freedom.

MAKING EFFICIENT USE OF ALL AVAILABLE DATA

This section presents two elementary applications of the BCR, first in the well-known case of forecast updating and second, in missing data estimation. We show that in both cases the optimal estimate can be obtained in the same way: we start with an initial set of forecasts obtained from historical data and we revise them by imposing the restriction that at some specific time points the forecasts must be equal to the corresponding observed values. By posing the problem in this framework we prove that the usual optimal estimators, already derived in the literature by other approaches, are also obtained with the BCR.

Let $X = (Z_1, \ldots, Z_T)$ be the historical data and $\mathbf{Z} = (Z_{T+1}, \ldots, Z_{T+H})'$ be the $H > 1$ future values to be forecasted with origin at time T. The vector of forecasts is $\hat{\mathbf{Z}}_p = (\hat{Z}_T(1), \ldots, \hat{Z}_T(H))' = E(\mathbf{Z} \mid X)$ and, when the series follows an ARIMA model with pure moving average (MA) representation $Z_t = \psi(B)a_t$, we know that

$$Z_{T+h} - E(Z_{T+h}|X) = \sum_{j=0}^{h-1} \psi_j a_{T+h-j} \quad \text{for } h = 1, \ \ldots, \ H$$

where $\psi(B) = (\psi_0 + \psi_1 B + \psi_2 B^2 + \ldots)$ with $\psi_0 = 1$, so that the ψ_j's are the MA weights of the model and $\{a_t\}$ is a zero-mean white-noise process with variance σ^2. This expression holds true for both stationary and non-stationary time series and can be rewritten in matrix notation as

$$\mathbf{Z} - \hat{\mathbf{Z}}_p = \mathbf{\Psi}_H \mathbf{a} \tag{8}$$

where $\mathbf{a} = (a_{T+1}, \ldots, a_{T+H})'$ and $\mathbf{\Psi}_H$ is the lower triangular matrix

$$\mathbf{\Psi}_H = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ \psi_1 & 1 & 0 & \ldots & 0 \\ \psi_2 & \psi_1 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \psi_{H-1} & \psi_{H-2} & \psi_{H-3} & \ldots & 1 \end{pmatrix}$$

We shall call $\mathbf{e} = \mathbf{\Psi}_H \mathbf{a}$ the forecast error and $\Sigma_e = \sigma^2 \mathbf{\Psi}_H \mathbf{\Psi}'_H$ will be its covariance matrix, which is given by

$$\Sigma_e = \begin{pmatrix} 1 & \psi_1 & \psi_2 & \ldots & \psi_{H-1} \\ \psi_1 & 1+\psi_1^2 & \psi_1 + \psi_1\psi_2 & \ldots & \Sigma_0^1 \psi_i \psi_{H-2+i} \\ \psi_2 & \psi_1 + \psi_1\psi_2 & 1+\psi_1^2+\psi_2^2 & \ldots & \Sigma_0^2 \psi_i \psi_{H-3+i} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \psi_{H-1} & \Sigma_0^1 \psi_i \psi_{H-2+i} & \Sigma_0^2 \psi_i \psi_{H-3+i} & \ldots & \Sigma_0^{H-1} \psi_i^2 \end{pmatrix} \tag{9}$$

Also $\Pi(\mathbf{Z} - \hat{\mathbf{Z}}_p) = \mathbf{a}$, where $\Pi = \Psi_H^{-1}$ is a lower triangular matrix with ones in the main diagonal, $-\pi_1$ in the first subdiagonal and so on, where the π_i's are the pure autoregressive (AR) coefficients of the process in the representation $\pi(B)Z_t = a_t$, with $\pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \ldots)$. Then $\Sigma_e^{-1} = \sigma^{-2}\Pi'\Pi$ represents the matrix of autocovariances of the inverse process $Z_t = \pi(B)a_t$ and it is called the inverse autocovariance matrix of the process.

Forecast updating

We consider here the problem of updating a vector of ARIMA forecasts, initially obtained with origin at time T. In this case, as soon as we observe the new observation, Z_{T+1}, the forecasts $\hat{Z}_T(2), \ldots, \hat{Z}_T(H)$ become suboptimal. We shall now prove that by introducing the restriction $\hat{Z}_T(1) = Z_{T+1}$ and revising the initial forecast vector via the BCR, we obtain the usual forecast updating equations. Taking $\hat{\mathbf{Z}}_p = (\hat{Z}_T(1), \ldots, \hat{Z}_T(H))'$, $\mathbf{Y} = Z_{T+1}$ and $C = (1, \mathbf{0}')$ with $\mathbf{0}$ a column vector of size $H-1$, and then calling $\hat{\mathbf{Z}} = (Z_{T+1}, \hat{Z}_{T+1}(1), \ldots, \hat{Z}_{T+1}(H-1))'$, $\Sigma_e = \sigma^2 \Psi_H \Psi'_H$ and $A = (1, \psi_1, \ldots, \psi_{H-1})'$, the BCR yields

$$\hat{\mathbf{Z}} = \hat{\mathbf{Z}}_p + A(Z_{T+1} - \hat{Z}_T(1))$$

which leads to the well-known equation (cf. Box and Jenkins, 1976, Ch. 5)

$$\hat{Z}_{T+1}(j) = \hat{Z}_T(j+1) + \psi_j(Z_{T+1} - \hat{Z}_T(1)) \quad \text{for } h = 1, \ldots, H-1$$

and

$$\Gamma = \sigma^2 \begin{pmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Psi}_{H-1}\boldsymbol{\Psi}'_{H-1} \end{pmatrix}$$

This matrix reflects the fact that the first term of $\hat{\mathbf{Z}}$ is observed and therefore has zero variance. In general we could use m $(1 \leqslant m < H)$ additional observations at once, if they arrive in a batch, or else we could update the forecasts recursively; the answer will be the same.

Next, the compatibility statistic in the situation $\mathbf{Y} = Z_{T+1}$ with $\{a_t\}$ Gaussian leads us to declare the new observation compatible with the historical record of the series, at the $100\alpha\%$ significance level, when

$$K = [Z_{T+1} - \hat{Z}_T(1)]^2/\sigma^2 < \chi_1^2(\alpha)$$

which is a standard forecasting test (see Box and Tiao, 1976). When the new data arrive in batches, the K-statistic can be used as a Cumulative Sum (CUSUM) test (See Inclán and Tiao, 1994).

Estimation of missing data

We now consider the problem of completing a univariate time series which has some missing values. We will show that the optimal estimation of the missing values can be obtained as follows. (1) Take the observation just before the first missing value as forecast origin and compute the vector of forecasts for all the sample period; (2) revise the forecasts using the BCR with the restriction that the new forecasts must be equal to the observed values. The revised forecasts for the missing value provide directly the optimal missing value estimates.

To prove this result, let us suppose that the series has $k \geqslant 1$ missing values (without any specific pattern) and include them in the vector $\mathbf{Z}_M = (Z_{T_1}, Z_{T_2}, \ldots, Z_{T_k})'$ in which $T_i < T_j$ if $i < j$. Let also $X = (Z_1, \ldots, Z_{T_1-1})$ be the historical data before the first missing value. This set will be used to generate the forecasts for the rest of the values. Let $\mathbf{Z}^* = (Z_{T_1}, \ldots, Z_T)'$ be the vector of observations to be forecasted and $\mathbf{Z}_G = (Z_{T_1+1}, Z_{T_2-1}, Z_{T_2+1}, \ldots, Z_{T_k-1}, Z_{T_k+1}, \ldots, Z_T)'$ be the vector of observed sample points after the first missing value. Then the random variable to be forecasted, $\mathbf{Z}$, can be written as $\mathbf{Z} = (\mathbf{Z}'_M, \mathbf{Z}'_G)' = \Omega\mathbf{Z}^*$ with Ω a permutation matrix that merely changes the chronological ordering of $\mathbf{Z}^*$. That is, Ω is a $T - T_1 + 1$ square matrix obtained from the identity matrix, in which the rows T_i, $i = 1, \ldots, k$, of the identity are placed as the first k rows of Ω. Thus $\hat{\mathbf{Z}}_p = E(\mathbf{Z}|X) = \Omega E(\mathbf{Z}^*|X)$ contains the ARIMA forecasts for the observations $Z_{T_1}, \ldots, Z_T$. The corresponding forecast error is $\mathbf{e} = \Omega\boldsymbol{\Psi}_{T-T_1+1}\mathbf{a}$, with covariance matrix

$$\Sigma_e = \sigma^2\Omega\boldsymbol{\Psi}_{T-T_1-1}\boldsymbol{\Psi}'_{T-T_1-1}\Omega' = \sigma^2\begin{pmatrix} \Sigma_M & \Sigma_{MG} \\ \Sigma'_{MG} & \Sigma_G \end{pmatrix} \tag{10}$$

where $\sigma^2\Sigma_M = \mathrm{Cov}[\mathbf{Z}_M - E(\mathbf{Z}_M|X)]$, $\sigma^2\Sigma_{MG} = E\{[\mathbf{Z}_M - E(\mathbf{Z}_M|X)][\mathbf{Z}_G - E(\mathbf{Z}_G|X)]'\}$ and $\sigma^2\Sigma_G = \mathrm{Cov}[\mathbf{Z}_G - E(\mathbf{Z}_G|X)]$.

We now revise the forecasts by introducing the restriction that the new forecasts for the observed values must be equal to the sample data. Then, by choosing $C = (0, I_{T-T_1-k+1})$ with 0 a $(T - T_1 - k + 1) \times k$ matrix, we obtain $\mathbf{Y} = \mathbf{Z}_G$ and the weighting matrix

$$A = \Sigma_e C'(C\Sigma_e C')^{-1} = \begin{pmatrix} \Sigma_{MG}\Sigma_G^{-1} \\ I_{T-T_1-k+1} \end{pmatrix}$$

Thus, by the BCR, the MMSELP of $\mathbf{Z}$ based on $E(\mathbf{Z} \mid X)$ and $\mathbf{Z}_G$ leads to $\hat{\mathbf{Z}}_G = \mathbf{Z}_G$, and the estimation of the missing values is given by

$$\hat{\mathbf{Z}}_M = E(\mathbf{Z}_M|X) + \Sigma_{MG}\Sigma_G^{-1}[\mathbf{Z}_G - E(\mathbf{Z}_G|X)] \tag{11}$$

with

$$\mathrm{Cov}(\hat{\mathbf{Z}}_M - \mathbf{Z}_M) = \sigma^2\Big(\Sigma_M - \Sigma_{MG}\Sigma_G^{-1}\,\Sigma'_{MG}) \tag{12}$$

To illustrate these equations, let us consider the case of a missing value, $k = 1$, at time $T_1 = h$ in an AR(1) process $(1 - \phi B)Z_t = a_t$. Then $Z_M = Z_h$, $X = (Z_1, \ldots, Z_{h-1})$, $\mathbf{Z}_G = (Z_{h-1}, \ldots, Z_T)'$, $\mathbf{Z}^* = (Z_h, \ldots, Z_T)'$, and $\Omega = I_{T-h+1}$. Also, by equation (10) $\Sigma_M = 1$, while the term Σ_{MG} Σ_G^{-1} can be obtained directly from the inverse of Σ_e. This matrix is given by $\sigma^{-2}\Pi'\Pi$, and in this case Π has ones in the diagonal, $-\phi$ in the next subdiagonal and zeros everywhere else. Thus the first row of this matrix is $\sigma^{-2}(1 + \phi^2, -\phi, 0, .., 0)$. On the other hand, using the standard formula for the inverse of the partitioned matrix (10) we know that this first row is given by $\sigma^{-2}(a, -a\Sigma_{MG}\Sigma_G^{-1})$, and so $a = (1 + \phi^2)$ and

$$\Sigma_{MG}\Sigma_G^{-1} = \left(\frac{\phi}{1+\phi^2},\ 0,\ \ldots,\ 0\right)$$

and the optimal estimator of the missing value is

$$\hat{Z}_h = \phi Z_{h-1} + \frac{\phi}{1+\phi^2}(Z_{h+1} - \phi^2 Z_{h-1}) = \frac{\phi}{1+\phi^2}(Z_{h+1} + Z_{h-1}) \tag{13}$$

This expression is the standard equation for estimating a missing value in an AR(1) process. When $\phi = 1$ the optimal estimator is just the average of the two contiguous observations, whereas when $\phi < 1$ this estimator shrinks the value towards the average of the process (zero in this case).

It is shown in the Appendix that the general expression for missing value estimation presented here is equivalent to the one obtained by Peña and Maravall (1991) using the inverse autocorrelation function. Alvarez, Delrieu and Jareño (1993) indicated the possibility of using restricted estimation to estimate missing values. From a practical standpoint, it is important to realize that the matrices involved in the missing value estimation may be relatively large, although, as shown in the previous example, the term $\Sigma_{MG}\Sigma_G^{-1}$ can be obtained from the matrix $\Sigma_e^{-1} = \sigma^{-2}\Pi'\Pi$, which is known given the process. This solution is equivalent to the one presented in the Appendix and can sometimes be computationally as efficient as the recursive solutions (see Gómez and Maravall, 1994; Gómez, Maravall and Peña, 1999).

The statistic to test for compatibility between $\mathbf{Z}_G$ and $E(\mathbf{Z}_G \mid X)$ becomes

$$K = [\mathbf{Z}_G - E(\mathbf{Z}_G|X)]'\Sigma_G^{-1}[\mathbf{Z}_G - E(\mathbf{Z}_G|X)]/\sigma^2 \sim \chi^2_{T-T_1-k+1}$$

and it compares all the data after the first missing value with the forecast of these data from the first period of complete data. It is interesting to realize that this CT can be considered as a generalization of the one proposed by Box and Tiao (1976) as an overall check of model validity because it checks whether the data posterior to the missing value is compatible with the previous ones. Rejection of the compatibility assumption may be expected when the missing values occur not just by chance, but are due to some exogenous intervention or structural changes in the time series which may affect several observations after the missing value.

RESTRICTED FORECASTING WITH BINDING CONSTRAINTS

The problem considered now is presented with two variants that may be deemed as relevant possibilities derived from compatibility testing.

Restricted forecasting without uncertainty in the restrictions

This case occurs when some restrictions to be imposed on the time series forecasts are known to be true in advance. For instance, we may consider imposing budget constraints, or else we may view this kind of application as a scenario (or what if) analysis. For instance, Guerrero (1989) faced the problem of forecasting the monthly Financing Granted by the Mexican Bank System when $\mathbf{Y}$, the total annual financing, was known in advance. Furthermore, the m restrictions in $\mathbf{Y} = C\mathbf{Z}$ are assumed to be linearly independent and coming from outside information (external to the model). On these conditions the BCR applies with $A = \mathbf{\Psi}_H\mathbf{\Psi}'_H C'(C\mathbf{\Psi}_H\mathbf{\Psi}'_H C')^{-1}$ and the corresponding CT statistic becomes

$$K = (\mathbf{Y} - C\hat{\mathbf{Z}}_p)'(C\mathbf{\Psi}_H\mathbf{\Psi}'_H C')^{-1}(\mathbf{Y} - C\hat{\mathbf{Z}}_p)/\sigma^2 \sim \chi^2_m$$

We emphasize the importance of compatibility testing, since rejecting this hypothesis may lead to different relevant formulations. For instance, we may assume that $\mathbf{Y}$ is true and that $\hat{\mathbf{Z}}_p$ is not a valid forecast, because a structural change is foreseen during the forecast horizon.

To shed some light into how the BCR works, let us assume that we want to forecast some quarterly expenses, with the restriction that their annual sum is equal to some budget constraint. Thus, we let $\hat{\mathbf{Z}}_p = (\hat{Z}_T(1), \ldots, \hat{Z}_T(4))'$, $C = \mathbf{1}' = (1, 1, 1, 1)$ and $Y = \Sigma_{h=1}^4 Z_{T+h}$. Then we have that

$$\hat{\mathbf{Z}} = \hat{\mathbf{Z}}_p + A[Y - \sum_{h=1}^{4} \hat{Z}_T(h)] \tag{14}$$

where $A = \mathbf{\Psi}_4\mathbf{\Psi}'_4\mathbf{1}(\mathbf{1}'\mathbf{\Psi}_4\mathbf{\Psi}'_4\mathbf{1})^{-1}$ is a vector of weights whose elements add up to one. In particular, when the series follows a white-noise process, $\mathbf{\Psi}_4 = I_4$ and $A = 1/4\mathbf{1}$, so the discrepancy between $\mathbf{Y}$ and the sum of the forecasts is distributed evenly among the quarters.

The square root of the CT statistic becomes in this case

$$K^{1/2} = |Y - \sum_{h=1}^{4} \hat{Z}_T(h)| \Big/ \left[\sigma \sqrt{\sum_{h=1}^{4} (\sum_{j=0}^{4-h} \psi_j)^2} \right] \tag{15}$$

which can be easily interpreted as a standardized normal test.

As an example suppose that we want to forecast four quarterly observations of a time series which follows the model $(1 - 0.5B)\ Z_t = a_t$, from origin at $T = 60$, where a_t are iid N(0,1) and $Z_{60} = 0.37$, satisfying the annual budget constraint given by $Y = C\mathbf{Z} = 3$ where $C = \mathbf{1}'$ and $\mathbf{Z} = (Z_{61}, \ldots, Z_{64})'$. First, we test for compatibility by calculating K as given by equation (15) with $\mathbf{Z}_p = (\hat{Z}_{60}(1), \ldots, \hat{Z}_{60}(4))'$ and

$$\mathbf{\Psi}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \phi & 1 & 0 & 0 \\ \phi^2 & \phi & 1 & 0 \\ \phi^3 & \phi^2 & \phi & 1 \end{pmatrix}$$

so that, as $\phi = 0.5$, $(\mathbf{1}'\mathbf{\Psi}_4\mathbf{\Psi}_4'\mathbf{1})^{-1} = 0.1017$ and $\hat{\mathbf{Z}}_p = (\phi Z_{60}, \ldots, \phi^4 Z_{60})' = (0.185, 0.092, 0.046, 0.023)'$, then $Y - \mathbf{1}'\hat{\mathbf{Z}}_p = 2.654$ and

$$K = (Y - \mathbf{1}'\hat{\mathbf{Z}}_p)'(\mathbf{1}'\mathbf{\Psi}_4\mathbf{\Psi}_4'\mathbf{1})^{-1}(Y - \mathbf{1}'\hat{\mathbf{Z}}_p)/\sigma^2 = 0.72$$

By comparing this figure with a Chi-square distribution with one degree of freedom we conclude that the restriction is compatible with the historical data at any sensible significance level.

The restricted forecasts are then obtained from equation (14), where $A = (0.191, 0.274, 0.289, 0.246)'$

$$\hat{\mathbf{Z}} = \hat{\mathbf{Z}}_p + A[Y - \sum_{h=1}^{4} \hat{Z}_{60}(h)] = (0.691,\ 0.818,\ 0.814,\ 0.677)'$$

It can be easily verified that the restriction is fulfilled exactly by these forecasts. Furthermore, the BCR also yields

$$\text{Cov}(\hat{\mathbf{Z}} - \mathbf{Z}) = \sigma^2(I_4 - A\mathbf{1}')\mathbf{\Psi}_4\mathbf{\Psi}_4' = \begin{pmatrix} 0.642 & -0.013 & -0.293 & -0.337 \\ -0.013 & 0.596 & -0.067 & -0.574 \\ -0.293 & -0.067 & 0.490 & 0.462 \\ -0.337 & -0.574 & 0.462 & 0.805 \end{pmatrix}$$

This matrix should be compared with

$$\text{Cov}(\hat{\mathbf{Z}}_p - \mathbf{Z}) = \sigma^2\mathbf{\Psi}_4\mathbf{\Psi}_4' = \begin{pmatrix} 1 & 0.5 & 0.25 & 0.125 \\ 0.5 & 1.25 & 0.625 & 0.016 \\ 0.25 & 0.625 & 1.313 & 0.656 \\ 0.125 & 0.016 & 0.656 & 1.328 \end{pmatrix}$$

which corresponds to the unrestricted ARIMA forecasts. We then reach the conclusion that a substantial gain in precision was obtained, particularly with the second and third forecasts, by incorporating the restrictions. It should be stressed, however, that the validity of these forecasts rely heavily on the compatibility assumption between restrictions and ARIMA forecasts. For instance, had the restriction been $Y = 5$, then the CT statistic would have taken the value $K = 3.25$, which becomes significant at the 10% level. In that situation, if the restriction is deemed valid, perhaps a structural change should be expected during the forecast horizon.

Change foreseen in the deterministic structure of the model

Let us now suppose that a structural change in the deterministic structure of the time series model is foreseen to occur during the forecast horizon of interest. This idea may come from subject matter considerations, for example when an intervention is anticipated. For instance, suppose we know that the time series measurement system is going to change due to a different weighting scheme of the time series components. This case may be considered as an *ex-ante* intervention analysis in which the whole effect of the intervention is presumably accounted for by way of some linear restrictions on the future values of the series. This kind of problems were considered by Guerrero (1991).

We start first with a formulation that allows us to take into account the intervention effects. That is, we assume that the future values of the series $\{Z_t^{(D)}\}$ will verify

$$Z_t^{(D)} = Z_t + D_t \tag{16}$$

with $\{Z_t\}$ the series without intervention effects and D_t a dynamic function of the intervention effects. Since in practice we usually have access only to one or two restrictions, we are forced to postulate at most the following first order linear dynamic model for D_t

$$(1 - \delta B)D_t = \omega S_t^{(\tau+1)} \tag{17}$$

where $S_t^{(\tau)}$ is a step function that takes on the value 1 when $t \geqslant \tau$ and is 0 otherwise, with τ the time point at which the intervention effects start, so that $T < \tau \leqslant T + h$. For simplicity we assume $\tau = T + 1$.

To employ the BCR let $\mathbf{Z} = (Z_{T+1}, \ldots, Z_{T+H})'$ be the vector of future values without intervention effects, $\mathbf{D} = (D_{T+1}, \ldots, D_{T+H})'$ be the deterministic effects and $\mathbf{Z}^{(D)} = (Z_{T+1}^{(D)}, \ldots, Z_{T+H}^{(D)})' = \mathbf{Z} + \mathbf{D}$ the unknown future observed values. Then, calling $\hat{\mathbf{Z}}_p = E(\mathbf{Z}|X) + \mathbf{D}$, and assuming that $\mathbf{Y} = C\mathbf{Z}^{(D)}$ imposes m linear independent restrictions, it follows that

$$\hat{\mathbf{Z}}^{(D)} = \hat{\mathbf{Z}}_p + \boldsymbol{\Psi}_H\boldsymbol{\Psi}'_H C'(C\boldsymbol{\Psi}_H\boldsymbol{\Psi}'_H C')^{-1}[\mathbf{Y} - C\hat{\mathbf{Z}}_p]$$

A difficulty with this equation is that $\hat{\mathbf{Z}}_p$ is assumed known, but in practice we can generate $E(\mathbf{Z} \mid X)$ and $\mathbf{D}$ has to be specified by solving

$$C\hat{\mathbf{D}} = \mathbf{Y} - CE(\mathbf{Z}|X) \tag{18}$$

which is assumed to be a system of consistent equations (i.e. any linear relationship existing among the rows of C also exists among the rows of $\mathbf{Y} - CE(\mathbf{Z} \mid X)$). Hence we get

$$\hat{\mathbf{Z}}^{(D)} = E(\mathbf{Z}|X) + \hat{\mathbf{D}} \tag{19}$$

On the other hand, the compatibility statistic is always zero in this case because $\hat{\mathbf{Z}}_p$ and $\mathbf{Y}$ are necessarily compatible by construction of $\hat{\mathbf{D}}$.

As an example, let us consider the problem of forecasting the monthly growth of a production index subjected to the constraint that the average growth during that year must be equal to Y. In this case it is assumed that an intervention, whose effects can be represented by equation (17) with $\delta = 0$ (implying a level change) will take place at the beginning of the year. The forecasts $E(\mathbf{Z} \mid X)$ are only used as a starting point and the main problem lies in obtaining $\hat{\mathbf{D}}$. Thus, calling $\mathbf{1} = (1, 1, \ldots, 1)'$ since in this case we know that $C = 1/12\mathbf{1}'$ and $\mathbf{D} = \omega\mathbf{1}$, then equation (18) implies

$$\hat{\omega} = Y - \frac{1}{12}\sum_{h=1}^{12}\hat{Z}_T(h)$$

This is not a surprising result, since it indicates estimating the intervention effect as the difference between the restriction and its corresponding unrestricted forecast.

ANALYSIS OF OUTLIERS IN TIME SERIES

Some of the developments of the previous sections are now applied to derive several important results in the context of outlier analysis. These examples may help to appreciate the potential usefulness of the BCR and the CT as employed in more specialized situations.

Detecting and measuring the effect of influential outliers

We consider first the single additive outlier case with known time of occurrence h. Additive outliers (AO) are considered for instance, by Tsay (1986), Chang, Tiao and Chen (1988) and Peña (1990). We assume that the observed series $\{Z_t^{(AO)}\}$ is generated as

$$Z_t^{(AO)} = Z_t + \omega_A P_t^{(h)}$$

where $P_t^{(h)}$ denotes the pulse function that takes on the value 1 when $t = h$ and is zero otherwise, and $\{Z_t\}$ follows an ARIMA model.

The AO can be posed as a missing value problem and solved as indicated previously by equations (11) and (12). From equation (A3) in the Appendix this predictor can also be written as

$$\hat{Z}_h = \mathbf{1}_h'\Pi'\Pi\mathbf{Z}_c$$

where in place of D we use $\mathbf{1}_h$ as a vector with one at position h and zero elsewhere, while $\mathbf{Z}_c$ has a zero at position h and the observed values otherwise. Now, calling ρ_i^I to the inverse autocorrelation coefficients and $\hat{Z}_h = E(Z_h \mid X^*)$, with $X^* = (X, \mathbf{Z}_G)$, this expression can be rewritten as

$$\hat{Z}_h = -\Sigma\rho_i^I(Z_{h-i} + Z_{h+i})$$

which is a well-known equation obtained by Brubacher and Wilson (1976) as the optimal interpolator of a time series. Then the optimal estimator of the outlier effect is given by

$$\hat{\omega}_A = Z_h^{(AO)} - E(Z_h|X^*)$$

with

$$\mathrm{Var}(\hat{\omega}_A) = \sigma^2(1 - \Sigma_{MG}\Sigma_G^{-1}\,\Sigma'_{MG})$$

Moreover, on the Gaussianity assumption for $\{a_t\}$ we obtain the following statistic for testing compatibility between $Z_h^{(AO)}$ and $E(Z_h \mid X^*)$:

$$K = \hat{\omega}_A^2/[\sigma^2(1 - \Sigma_{MG}\Sigma_G^{-1}\,\Sigma'_{MG})] \sim \chi_1^2$$

Noting that $(1 - \Sigma_{MG}\Sigma_G^{-1}\Sigma'_{MG})^{-1}$ is the (1,1) element of the matrix $\Sigma_e^{-1} = \sigma^{-2}\Pi'\Pi$, it is clear that

$$1 - \Sigma_{MG}\Sigma_G^{-1}\,\Sigma'_{MG} = \sum_{i=0}^{T-h} \pi_i^2$$

Therefore, the CT statistic can be written as $\hat{\omega}_A^2/(\sigma^2\Sigma\pi_i^2)$, which is the likelihood ratio test proposed by Chang, Tiao and Chen (1988) to test for additive outliers. This statistic is also related to the influence measures derived by Peña (1990).

Reallocation outliers

A situation considered and illustrated by Wu, Hosking and Ravishanker (1993) consists of restricting a block of consecutive observations affected by outliers to produce the same sum as if no outliers were present. So let us consider the multiple additive outlier formulation

$$Z_t^{(RO)} = Z_t + \sum_{i=0}^{m} \lambda_i P_t^{(c+i)} \tag{20}$$

where $\{Z_t^{(RO)}\}$ is the series with Reallocation Outliers (RO), $P_t^{(c+i)}$ is the pulse function defined as before and λ_i is the effect associated with the outlier occurring at time $c+i$. The values of c and m (timing and duration of reallocation) are assumed known, although Wu, Hosking and Ravishanker (1993) also addressed the case in which c and/or m are unknown.

Now let us define $\mathbf{Z}^{(RO)} = (Z_c^{(RO)}, Z_{c+1}^{(RO)}, \ldots, Z_{c+m}^{(RO)})'$ and $\mathbf{Z}_M = (Z_c, Z_{c+1}, \ldots, Z_{c+m})'$ as the vectors of observations with and without outlier effects. Similarly let $X = (Z_1, \ldots, Z_{c-1})$ and $\mathbf{Z}_G = (Z_{c+m+1}, \ldots, Z_T)'$ be the vectors of past and future observations with respect to $Z^{(RO)}$, and $X^* = (X, \mathbf{Z}_G)$. Then, calling $\hat{\mathbf{Z}}_p = E(\mathbf{Z}_M \mid X)$ and if we consider that $Y = \mathbf{1}'\mathbf{Z}^{(RO)}$ is given, with $\mathbf{1}$ an m-vector of ones, the BCR yields

$$\hat{\mathbf{Z}}_M = \hat{\mathbf{Z}}_p + \Sigma_e\mathbf{1}(\mathbf{1}'\Sigma_e\mathbf{1})^{-1}\mathbf{1}'(\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_p)$$

with

$$\Gamma = \Sigma_e - \Sigma_e\mathbf{1}(\mathbf{1}'\Sigma_e\mathbf{1})^{-1}\mathbf{1}'\Sigma_e \tag{21}$$

Next, from equation (20) we have $\mathbf{Z}^{(RO)} = \mathbf{Z} + \boldsymbol{\lambda}$ with $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m)'$, so that

$$\hat{\boldsymbol{\lambda}} = \mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_M = (\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_p) - \Sigma_e\mathbf{1}(\mathbf{1}'\Sigma_e\mathbf{1})^{-1}\mathbf{1}'(\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_p) \tag{22}$$

and $\mathrm{Cov}(\hat{\boldsymbol{\lambda}}) = \Gamma$. Furthermore, the CT statistic (7) for testing compatibility between $Y = \mathbf{1}'\mathbf{Z}^{(RO)}$ and $\hat{\mathbf{Z}}_p$, on the normality assumption for $\mathbf{e}$, is given by

$$K = (\mathbf{1}'\mathbf{Z}^{(RO)} - \mathbf{1}'\hat{\mathbf{Z}}_p)'(\mathbf{1}'\Sigma_e\mathbf{1})^{-1}(\mathbf{1}'\mathbf{Z}^{(RO)} - \mathbf{1}'\hat{\mathbf{Z}}_p) \\ = [\mathbf{1}'(\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_p)]^2/(\mathbf{1}'\Sigma_e\mathbf{1}) \sim \chi_1^2 \tag{23}$$

Expressions (21)–(23) are easily seen to be the same as those deduced by Wu, Hosking and Ravishanker (1993), by recognizing that in their notation $\sigma^2 A^{-1}$ and $A^{-1}b$ are Σ_e and $\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_p$ in ours. In particular expression (23) is useful for testing whether the additive outliers are reallocation or not. Moreover, they also proposed a statistic for testing the null hypothesis of no outlier effects versus influential outliers, which in our context becomes a test for compatibility between $\mathbf{Z}^{(RO)}$ and $\hat{\mathbf{Z}}_M$, that is.

$$K^* = (\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_M)'\Sigma_e^{-1}(\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_M) \sim \chi_m^2 \tag{24}$$

Let us consider as an example of this situation the case in which a promotional campaign launched by a company increased its sales during one quarter, but it is feared that the increase in sales at time c (the previous quarter) is compensated by a decrease in the sales of the current quarter. That is, we assume that the campaign took place at time $c = 59$ and the duration of the reallocation is $m = 1$, so that $\lambda_0 = -\lambda_1$, $\mathbf{Z}^{(RO)} = (Z_{59} + \lambda_0, Z_{60} + \lambda_1)'$, $\mathbf{Z}_M = (Z_{59}, Z_{60})'$ and $X = (Z_1, \ldots, Z_{58})$, whereas no additional data $\mathbf{Z}_G$ exist. We assume the same simulated data used in the previous example with the model $(1 - 0.5B)\, Z_t = a_t$ where a_t are iid N(0,1). In this case the simulated values are $Z_{58} = -0.964$, $Z_{59} = -0.260$, $Z_{60} = 0.369$ and $\lambda_0 = 4$. The problem is to estimate $\boldsymbol{\lambda} = (\lambda_0, \lambda_1)'$ on the basis of the observed values affected by the promotional campaign $\mathbf{Z}^{(RO)} = (3.740, -3.631)'$ and the historical data X, producing the forecasts $\hat{\mathbf{Z}}_p = E(\mathbf{Z}_M|X) = (\phi Z_{58}, \phi^2 Z_{58})'$ with

$$\Sigma_e = \boldsymbol{\sigma}^2\boldsymbol{\Psi}_2\Psi_2' = \sigma^2\begin{pmatrix} 1 & \phi \\ \phi & 1+\phi^2 \end{pmatrix} \quad \text{where } \phi = 0.5$$

Then we obtain

$$\hat{\mathbf{Z}}_M = \hat{\mathbf{Z}}_p + A(\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_p) = (-0.320,\ -0.053)'$$

since

$$A = \Sigma_e\mathbf{1}(\mathbf{1}'\Sigma_e\mathbf{1})^{-1}\mathbf{1}' = \begin{pmatrix} 0.462 & 0.462 \\ 0.538 & 0.538 \end{pmatrix}$$

A relevant question in this context is whether the promotion had a significant effect at all. To answer it we can apply expression (24) in such a way that

$$K^* = (\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_M)'\Sigma_e^{-1}(\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_M) = 47.93$$

leads to rejection of the null hypothesis of non-significant effects in favour of the alternative of influential outliers. Hence we may proceed to calculate

$$\hat{\boldsymbol{\lambda}} = \mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_M = (4.060,\ -3.578)'$$

with

$$\mathrm{Cov}(\hat{\boldsymbol{\lambda}}) = (I_2 - A)\Sigma_e = \begin{pmatrix} 0.308 & -0.308 \\ -0.308 & 0.308 \end{pmatrix}$$

so that 95% confidence intervals for λ_0 and λ_1 are, respectively, (2.972, 5.148) and (−4.666, −2.490), which obviously cover the true values.

Another interesting question to be considered now is whether the outliers (that we already know have significant effects) are in fact reallocation or not. The answer in this case is fairly obvious from the confidence intervals, but it is also supported by the CT statistic

$$K = [\mathbf{1}'(\mathbf{Z}^{(RO)} - \hat{\mathbf{Z}}_p)]^2/(\mathbf{1}\Sigma_e\mathbf{1}) = 20.16$$

which, when compared against a χ_1^2 distribution, indicates rejection of the null hypothesis of no reallocation.

TEMPORAL DISAGGREGATION OF TIME SERIES

The problem of temporal disaggregating a time series is that of estimating an unobserved random vector $\mathbf{Z} = (Z_1, \ldots, Z_{mn})'$ on the basis of knowing some linear aggregates $Y_i = \Sigma_{j=1}^n c_i Z_{n(i-1)+j}$, with $i = 1, \ldots, m$. Here, n denotes the intraperiod frequency of observation (i.e. if $\{Y_i\}$ is observed annually and $\{Z_t\}$ is a monthly series, $n = 12$) m is the number of whole-period observations and $\mathbf{c} = (c_1, \ldots, c_n) \neq \mathbf{0}$. Some usual forms of $\mathbf{c}$ are: $\mathbf{c} = (0, 0, \ldots, 0, 1)'$ for interpolating a stock series, $\mathbf{c} = \mathbf{1}'$ for distributing a flow series and $\mathbf{c} = 1/n\mathbf{1}'$ for distributing an index series. Here we can appreciate again that the BCR provides a starting solution on which the analyst may elaborate to obtain a final working solution, depending on different assumptions.

Temporal disaggregation without auxiliary data

Let us assume that $\{Z_t\}$ admits an ARIMA representation, and expression (8) holds true, with $\mathbf{a} = \mathbf{a}_Z = (a_{Z,1}, \ldots, a_{Z,mn})'$ such that $E(\mathbf{a}_Z \mid X) = \mathbf{0}$ and $E(\mathbf{a}_Z\mathbf{a}_Z'|X) = \sigma_Z^2 I_{mn}$, where $X = (\ldots, Z_{-1}, Z_0)$ denotes the infinite past of the series. In fact, for an ARIMA model to be reasonable in this situation we assume that the process started at a finite time point with fixed initial values. The vector $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ can be written in the form $\mathbf{Y} = C\mathbf{Z}$ by defining $C = I_n \otimes c'$, where $\otimes$ denotes Kronecker product. Then the problem is posed as in the BCR, so that the MMSELP of $\mathbf{Z}$, when $E(\mathbf{Z} \mid X)$, $\mathbf{Y}$ and $\sigma_Z^2 \boldsymbol{\Psi}_{mn}\boldsymbol{\Psi}_{mn}'$ are known, is of the form (3). However, in practice such a formula is useless because X is unknown, as well as $E(\mathbf{Z} \mid X)$ and $\sigma^2\boldsymbol{\Psi}_{mn}\boldsymbol{\Psi}_{mn}'$. Some approaches that have been followed to overcome these difficulties are the following.

(1) Assume *a priori* that $E(\mathbf{Z} \mid X)$ and $\boldsymbol{\Psi}_{mn}\boldsymbol{\Psi}_{mn}'$ have some simple structures, say $[I_{mn} - \boldsymbol{\Psi}_{mn}\boldsymbol{\Psi}_{mn}'C'(C\boldsymbol{\Psi}_{mn}\boldsymbol{\Psi}_{mn}'C')^{-1}C]E(\mathbf{Z}|X) = \mathbf{0}$ and $\boldsymbol{\Psi}_{mn}$ is derived as if an integrated process of order one or two were an adequate representation for $\{Z_t\}$. Then calculate

$$\hat{\mathbf{Z}} = \boldsymbol{\Psi}_{mn}\boldsymbol{\Psi}_{mn}'C'(C\boldsymbol{\Psi}_{mn}\boldsymbol{\Psi}_{mn}'C')^{-1}\mathbf{Y} \tag{25}$$

Such a solution was essentially proposed by Lisman and Sandee (1964) and by Boot, Feibes and Lisman 1967) when $n = 4$, and by Cohen, Müller and Padberg (1971) for arbitrary n.

(2) Assume that the model for $\{Z_t\}$ can be somehow known to the analyst, perhaps by assuming that some disaggregated observations exist. This allows the specification of the matrix $\mathbf{\Psi}_{mn}$ *a priori*, so that we can calculate $\hat{\mathbf{Z}}$ if $E(\mathbf{Z} \mid X)$ is also known. That was the approach of Harvey and Pierse (1984).
(3) Derive the disaggregate ARMA model of the stationary series $\{(1-B)^d Z_t\}$ from that of $\{(1-B)^d Y_i\}$ using the theoretical relationship that links those two series. Thus, obtain the autocovariance matrix of $\{(1-B)^d Z_t\}$ and use it in an expression similar to (25). Then obtain the predictor of $\mathbf{Z}$ from the previous one, by applying a linear operator that essentially services to transform $((1-B)^d \hat{\mathbf{Z}}_{d+1}, \ldots, (1-B)^d \hat{\mathbf{Z}}_{mn})'$ into $\hat{\mathbf{Z}} = (\hat{\mathbf{Z}}_1, \ldots, \hat{\mathbf{Z}}_{mn})'$. This solution was developed by Wei and Stram (1990); see also Stram and Wei (1986).

As an example we now assume that the only data available consists of $n = 3$ annual figures of $\{Y_i\}$, and we want to obtain the 12 unknown data points of the quarterly series $\{Z_t\}$, given that we assume that the model for quarterly observations is $(1 - 0.5B)\, Z_t = a_t$ where a_t are iid N(0,1), so that we can apply Harvey and Pierse's (1984) solution. The corresponding series are linked through the linear expression $\mathbf{Y} = C\mathbf{Z}$ where $C = I_3 \otimes c'$ with $\mathbf{c} = (1, 1, 1, 1)'$, in such a way that $Y_i = \Sigma_{j=1}^4 c_i Z_{4(i-1)+j}$, for $i = 1, 2, 3$. The unobservable data are assumed to be

t	1	2	3	4	5	6	7	8	9	10	11	12
Z_t	−0.235	0.958	1.034	0.551	−0.238	−2.289	−2.477	−1.585	−1.841	−1.891	1.329	0.528

while the annual observations are $\mathbf{Y} = (2.308, -6.589, -1.875)'$. As

$$\mathbf{\Psi}_{12} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \phi & 1 & \ldots & 0 \\ \phi^2 & \phi & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \phi^{11} & \phi^{10} & \ldots & 1 \end{pmatrix} \quad \text{where } \phi = 0.5$$

In that case, expression (25) produces the estimate

$$\hat{\mathbf{Z}} = (0.706,\ 0.844,\ 0.669,\ 0.088,\ -1.213,\ -1.789, -1.914,\\ -1.673,\ -0.878,\ -0.506,\ -0.327,\ -0.163)'$$

whose elements are not only close to the true quarterly Z values, but also add up to the annual data, as they should.

Temporal disaggregation on the basis of a preliminary series

The problem now is the same as before, except that X will no longer be considered as the infinite past of the series. Rather $\hat{\mathbf{Z}}_p = E(\mathbf{Z} \mid X)$ will be considered a preliminary estimate, derived perhaps from a set of information on auxiliary variables X which help to explain the behaviour of $\{Z_t\}$ or else $\hat{\mathbf{Z}}_p$ is observed directly. In both cases, a preliminary vector $\hat{\mathbf{Z}}_p$ is known and it will be assumed to be the MMSELP of $\mathbf{Z}$. Now let us also assume that $\mathbf{a} = (a_1, \ldots, a_{mn})'$ such that $E(\mathbf{a} \mid \hat{\mathbf{Z}}_p) = \mathbf{0}$ and $E(\mathbf{aa}' \mid \hat{\mathbf{Z}}_p) = \sigma^2 P$, where P is a positive definite matrix. This assumption may be justified by assuming that $\{Z_t\}$ and $\{\hat{Z}_{p,t}\}$ follow the same ARIMA model, with the same AR and MA parameters, but different white noise-generating processes. Such an assumption makes sense when $\hat{\mathbf{Z}}_p$ is indeed a preliminary estimate of $\mathbf{Z}$ and it allows derivation of the matrix $\boldsymbol{\Psi}_{mn}$ from the observed data. It should also be noted that the usual specification of $\mathbf{Z}$ as $\hat{\mathbf{Z}}_p$ plus noise is not assumed here, because it would imply a different autocorrelation structure for $\{Z_t\}$ and $\{\hat{Z}_{p,t}\}$.

Thus, if $\hat{\mathbf{Z}}_p$, $\Sigma_e = \sigma^2 \boldsymbol{\Psi}_{mn} \mathrm{P} \boldsymbol{\Psi}'_{mn}$ and $\mathbf{Y}$ are known, we obtain by the BCR

$$\hat{\mathbf{Z}} = \hat{\mathbf{Z}}_p + \boldsymbol{\Psi}_{mn} P \boldsymbol{\Psi}'_{mn} C'(C \boldsymbol{\Psi}_{mn} P \boldsymbol{\Psi}'_{mn} C')^{-1} (\mathbf{Y} - C\hat{\mathbf{Z}}_p) \tag{26}$$

and

$$\Gamma = \sigma^2 [I_{mn} - \boldsymbol{\Psi}_{mn} P \boldsymbol{\Psi}'_{mn} C'(C \boldsymbol{\Psi}_{mn} P \boldsymbol{\Psi}'_{mn} C')^{-1} C] \boldsymbol{\Psi}_{mn} P \boldsymbol{\Psi}'_{mn} \tag{27}$$

Guerrero (1990) showed that the corresponding expressions proposed by Denton (1971) are particular cases of (26) and (27) when $\hat{\mathbf{Z}}_p$ is directly observed. Similarly, when comparing Chow and Lin's (1971) solution with (26), it is clear that they focused their attention on simultaneously estimating $\mathbf{Z}$ and the regression parameters linking the auxiliary information X with $\mathbf{Z}$. By doing so they did not pay much attention to the potential autocorrelation structure in the regression errors. In addition, the covariance matrix of their solution is (28) plus another term that can be related to the fact that only X (not $\hat{\mathbf{Z}}_p$) was assumed to be given.

To apply (26) and (27) in practice, we require knowing not just $\boldsymbol{\Psi}_{mn}$ (which is obtainable from the ARIMA model for $\{\hat{Z}_{p,t}\}$), but P as well. Guerrero (1990) suggested a feasible solution based on a two-step procedure, akin to using estimated generalized least squares.

CONCLUSIONS

We have shown that the BCR for combining information from two different sources is a very useful tool for solving time series problems. Such a rule produces in fact a weighted average of the two predictors coming from both sources of information. Its optimality is easily revealed by the corresponding MSE matrix which is not only minimum for the class of linear and unbiased predictors considered but shows that using the extra-model information reduces the original variability in the model predictor.

Realizing that many statistical procedures are derived by combining information is important from a unifying point of view. Besides we advocate the use of the CT in order to appreciate whether the combination makes sense or not. Some of these tests have already appeared in the time series literature, associated mainly with likelihood-based inferences. Both the BCR and the CT are very simple statistical tools that have helped (and will surely keep helping) time series analysts to solve other practical problems.

APPENDIX

The relationship between additive outliers and missing observations leads to estimating the missing values by inserting zeros at the missing points and adding dummy variables to the model for each of these zero values to represent an additive outlier. Calling D the matrix of dummy variables, it can be shown that the MMSELP of $\mathbf{Z}_M$ is given by

$$\hat{\mathbf{Z}}_M = -(D'\Sigma_e^{-1}D)^{-1}D'\Sigma_e^{-1}\mathbf{Z}_c \tag{A1}$$

where $\mathbf{Z}_c$ denotes the completed series with zeros at the missing points. The MSE matrix of this estimator is then given by

$$\mathrm{Cov}(\hat{\mathbf{Z}}_M - \mathbf{Z}_M) = \sigma^2 D'\Sigma_e^{-1}D \tag{A2}$$

To show that expressions (A1) and (21) are equivalent, let us first recall that $\Sigma_e^{-1} = \sigma^{-2}\Pi'\Pi$, so that

$$\hat{\mathbf{Z}}_M = (D'\Pi'\Pi D)^{-1}D'\Pi'\Pi\mathbf{Z}_c \tag{A3}$$

Now, for simplicity let us suppose that the k missing values are consecutive and located at the time points $h, h+1, \ldots, h+k-1$, so that $D = (0, I_k, 0)'$. Then let us partition Π as follows:

$$\Pi = \begin{pmatrix} \Pi_1 & 0 & 0 \\ \Pi_2 & \Pi_4 & 0 \\ \Pi_3 & \Pi_5 & \Pi_6 \end{pmatrix} \tag{A4}$$

with Π_1, Π_4 and Π_6 square matrices of dimensions $h-1$, k and $T-h-k+1$ respectively, while Π_2, Π_3 and Π_5 are rectangular arrays of appropriate dimensions. Similarly, let $\mathbf{Z}_c = (X, 0, \mathbf{Z}_G)'$, so that (A1) can be rewritten as

$$\hat{\mathbf{Z}}_M = -(\Pi_4'\Pi_4 + \Pi_5'\Pi_5)^{-1}[(\Pi_4'\Pi_2 + \Pi_5'\Pi_3)X + \Pi_5'\Pi_6\mathbf{Z}_G] \tag{A5}$$

Then, inverting (10) by blocks taking into account that $\Sigma_e^{-1} = \sigma^{-2}\Pi'\Pi$ and using (A4) we obtain $\Sigma_{MG}\Sigma_G^{-1} = -(\Pi_4'\Pi_4 + \Pi_5'\Pi_5)^{-1}\Pi_5'\Pi_6$. On the other hand, calling $\hat{\mathbf{Z}}_* = (E(\mathbf{Z}_M \mid X)', E(\mathbf{Z}_G \mid X)')'$ and $\mathbf{Z}_H' = (X, 0', 0')$ we have

$$\hat{\mathbf{Z}}_* = -(D_H'\Pi'\Pi D_H)^{-1}D_H'\Pi'\Pi\mathbf{Z}_H \tag{A6}$$

where

$$D_H = \begin{pmatrix} 0 & I_k & 0 \\ 0 & 0 & I_{T-T-k+1} \end{pmatrix} \tag{A7}$$

After some algebraic manipulations and inserting expressions (A6) and (A7) into (11), we obtain (A5).

ACKNOWLEDGEMENTS

The first author's participation was made possible through a Professorship in Time Series Analysis and Forecasting in Econometrics provided by Asociacion Mexicana de Cultura. The second author has been supported by Catedra BBV de Calidad and by DGES under grant PB96-0111. The authors gratefully acknowledge the detailed comments and suggestions of two anonymous referees, an associate editor and one of the editors of this journal. They led to significant improvements in the presentation of our work.

REFERENCES

Alvarez, L. J., Delrieu, J. C. and Jareño, J. 'Tratamiento de predicciones conflictivas: Empleo eficiente de información extramuestral', *Estadística Española*, **35** (1993), 439–461.

Boot, J. C. G., Feibes, W. and Lisman, J. H. C. 'Further methods of derivation of quarterly figures from annual data', *Appl. Statist.*, **16** (1967), 65–75.

Box, G. E. P. and Jenkins, G. M. *Time Series Analysis, Forecasting and Control*, rev. edn, San Francisco: Holden-Day, 1976.

Box, G. E. P. and Tiao, G. C. 'Comparison of forecast and actuality', *Appl. Statist.*, **25** (1976), 195–200.

Brubacher, S. R. and Wilson, G. T. 'Interpolating time series with application to the estimation of holiday effects on electricity demand', *Appl. Statist.*, **25** (1976), 107–116.

Chang, I., Tiao, G. C. and Chen, C. 'Estimation of time series parameters in the presence of outliers', *Technometrics*, **30** (1988), 193–204.

Cholette, P. A. and Dagum, E. B. 'Benchmarking time series with autocorrelated survey errors', *Int. Statist. Rev.*, **62** (1994), 365–377.

Chow, G. C. and Lin, A. 'Best linear interpolation, distribution and extrapolation of time series by related series', *Rev. Econ. Statist.*, **53** (1971), 372–375.

Cohen, K. J., Müller, W. and Padberg, M. W. 'Autoregressive approaches to disaggregation of time series data', *Appl. Statist.*, **20** (1971), 119–129.

de Alba, E. 'Temporal disaggregation of time series: a Bayesian analysis', *J. Business and Economic Statistics*, **6** (1988), 197–206.

de Alba, E. 'Constrained forecasting in autoregressive time series models: a Bayesian Analysis', *Int. J. Forecasting*, **9** (1993), 95–108.

Denton, F. T. 'Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization', *J. Am. Statist. Assoc.*, **66** (1971), 99–102.

Draper, D., Gaver, D. P. Jr, Goel, P. K., Greenhouse, J. B., Hedges, L. V., Morris, C. N., Tucker, J. R. and Waternaux, C. M. *Combining Information. Statistical Issues and Opportunities for Research*, Washington, DC: National Academy Press, 1992.

Gómez, V. and Maravall, A. 'Estimation, prediction and interpolation for nonstationary series with the Kalman filter', *J. Am. Statist. Assoc.*, **89** (1994), 611–624.

Gómez, V., Maravall, A. and Peña, D. 'Missing observations in ARIMA models: skipping approach versus additive outlier approach', *J. Econometrics*, **88** (1999), 341–363.

Guerrero, V. M. 'Optimal conditional ARIMA forecasts', *J. Forecasting*, **8** (1989), 215–229.

Guerrero, V. M. 'Temporal disaggregation of time series. An ARIMA-based approach', *Int. Statist. Rev.*, **58** (1990), 29–46.

Guerrero, V. M. 'ARIMA forecasts with restrictions derived from a structural change', *Int. J. Forecasting*, **7** (1991), 339–347.

Harvey, A. C. and Pierse, R. G. 'Estimating missing observations in economic time series', *J. Am. Statist. Assoc.*, **79** (1984), 125–131.

Hedges, L. V. and Olkin, I. *Statistical Methods for Meta-Analysis*, New York: Academic Press, 1985.

Hillmer, S. C. and Trabelsi, A. 'Benchmarking of economic time series', *J. Amer. Statist. Assoc.*, **82** (1987), 1064–1071.

Inclán, C. and Tiao, G. C. 'Use of cumulative sums of squares for retrospective detection of changes of variance', *J. Amer. Statist. Assoc.*, **89** (1994), 913–923.

Jones, R. 'Maximum likelihood fitting of ARMA models to time series with missing observations', *Technometrics*, **22** (1980), 389–395.

Lisman, J. H. C. and Sandee, J. 'Derivation of quarterly figures from annual data', *Appl. Statist.*, **13** (1964), 87–90.

Min, Ch. and Zellner, A. 'Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates', *J. Econometrics*, **56** (1993), 89–118.

Palm, F. C. and Zellner, A. 'To combine or not to combine? Issues of combining forecasts', *J. Forecasting*, **11** (1992), 687–701.
Peña, D. 'Influential observations in time series', *J. Bus. Econ. Statist.*, **8** (1990), 235–241.
Peña, D. 'Combining information in statistical modeling', *The American Statistician*, **51** (1997), 326–332.
Peña, D. and Maravall, A. 'Missing observations, additive outliers and inverse auto-correlation function', *Comm. Statist. (Theory and Methods)*, **A-20** (1991), 3175–3186.
Sprott, D. A. 'Gauss's contributions to statistics', *Historia Matematica*, **5** (1978), 183–203.
Stram, D. O. and Wei, W. W. S. 'A methodological note on the disaggregation of time series totals', *J. Time Ser. Anal.*, **7** (1986), 293–302.
Tsay, R. S. 'Time series model specification in the presence of outliers', *J. Amer. Statist. Assoc.*, **81** (1986), 132–141.
Wei, W. W. S. and Stram, D. O. 'Disaggregation of time series models', *J. Roy. Statist. Soc.*, **B-52** (1990), 453–467.
Wu, L. S. Y., Hosking, J. R. M. and Ravishanker, N. 'Reallocation outliers in time series', *Appl. Statist.*, **42** (1993), 301–313.
Young, P. *Recursive Estimation and Time-Series Analysis*, Berlin: Springer-Verlag, 1984.

Authors' biographies:
Victor M. Guerrero is Chairman of the Department of Statistics at ITAM. He has published articles in *JBES*, *Management Science*, *International Journal of Forecasting*, *Biometrika* and several other journals. His research interests include econometrics, time-series analysis and forecasting.

Daniel Peña is Professor of Statistics at the Universidad Carlos III de Madrid. He has held visiting positions at the University of Wisconsin-Madison and the University of Chicago. His papers have appeared in *JASA*, *Technometrics*, *JBES*, *Biometrika* and *JRSS-B*, among other journals. His research interests are time series, Bayesian statistics and robust methods.

Authors' addresses:
Victor M. Guerrero, Departamento de Estadistica, Instituto Tecnológico Autónomo de México, México 01000, D. F. MEXICO.

Daniel Peña, Departmento de Estadística y Econometría, Universidad Carlos III de Madrid, Getafe, Madrid 28903, Spain.

