Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE DIRECT

Journal of Statistical Planning and Inference 116 (2003) 249–276

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Combining multiple time series predictors: a useful inferential procedure

Víctor M. Guerrero[a,*], Daniel Peña[b]

[a] Departamento de Estadística, Instituto Tecnológico Autónomo de México (ITAM), 01000 D.F. México, Mexico

[b] Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Getafe, Madrid, Spain

Received 20 January 2001; accepted 5 March 2002

Abstract

We present a general result that allows us to combine data from two different sources of information in order to improve the efficiency of predictors within the context of multiple time series analysis. Such a result is derived from generalized least squares and is given as a combining rule that takes into account the possibility of correlation between forecasts and bias in one of them. We then specialize that result to situations in which the predictors are unbiased and uncorrelated. Afterwards we propose measuring precision shares and testing for compatibility in order for the combination to make sense. Several applications of the combining rule are presented according to the nature of the linear constraints imposed by one of the data sources. When the constraints are binding we consider the case of restricted forecasts with exact linear restrictions, deterministic changes in the model structure and partial information on some variables. When the constraints are stochastic we study forecast combinations that include expert judgments and benchmarking. Thus, the connections among different standard techniques are emphasized by the combining rule and its companion compatibility test. An empirical example illustrates the usefulness of this inferential procedure in practice. © 2002 Elsevier Science B.V. All rights reserved.

MSC: 62M10; 62M20; 62P20

Keywords: Benchmarking; Compatibility testing; Forecast combination; Generalized least squares; Restricted prediction

* Corresponding author. Tel.: (+52)55-5628-4084; fax: (+52)55-5628-4086.
E-mail address: guerrero@itam.mx (V.M. Guerrero).

PII: S0378-3758(02)00186-6

1. Introduction

Multiple time series models are some of the most widely used tools employed to represent dynamic random phenomena. We assume that a k-dimensional vector of time series $\mathbf{Z}_t = (Z_{1t}, \ldots, Z_{kt})'$ is observed for $t = 1, \ldots, T$, and we are interested in predicting a set of future values of the vector time series, say $\mathbf{Z}$. To that end, a multivariate linear time series model is first built to provide a vector of forecasts. It could be a vector auto-regressive and moving average (VARMA) model (see Tiao and Box, 1981; Lütkepohl, 1991; Reinsel, 1993) a simultaneous equation model (see Judge et al., 1980) or a state space model (see Aoki, 1990; Shumway and Stoffer, 2000). We allow the forecasts generated from the model to have some unknown bias and we shall suppose that the model as well as its parameters are given, thus we do not touch upon such issues as model specification or model estimation. In order to predict $\mathbf{Z}$ we also assume that an additional source of information provides another vector of forecasts, $\mathbf{Y}$, for a set of linear combinations of the random vector $\mathbf{Z}$. We want to take this information into account in order to improve upon the original forecast vector.

The problem considered is very general and includes as particular cases many time series problems, such as updating forecasts when new information is available, combining forecasts, interpolation and missing value estimation, analysis of influential observations and outliers (including reallocation outliers), temporal disaggregation and benchmarking time series. We shall derive a general solution to this problem and show that it yields as special cases many interesting and, in many cases, well-known results. The idea of combining estimators or predictors is not new and many well-established methods rely on it, as we shall show in what follows. However, to the best of our knowledge the combining rule that we present in the next section generalizes many particular results that have appeared in the time series literature. Thus, the main goal of this paper is to provide a unifying view of several apparently different problems, in terms of their corresponding solutions.

This paper is organized as follows. Section 2 introduces the notation and presents the general rule applicable to a multiple time series for which a model provides the minimum mean square error linear predictor (MMSELP) and some extra-model information is assumed to be given by means of a stochastic linear constraint. These two sources of information are exploited to form a combined optimal predictor by way of generalized least squares (GLS). Such a predictor is then particularized to several situations depending on whether bias can be reasonably assumed to be present or not, and whether the two forecasts to be combined are uncorrelated or not. Some further analytical tools here derived and related to the combining rule are: (i) a measure of precision share attributable to each source of information and (ii) a test for compatibility of the predictors to be combined. These tools are particularly relevant to make inferences in practical applications. In fact, they are useful to indicate how much is to be gained by combining and to make sure that the combination makes sense from an empirical point of view.

Section 3 shows how the combining rule can be applied when the constraint is binding. Some problems that can be analyzed under this framework are restricted forecasting with or without incompatible sources of information, deterministic changes in the model structure and taking into account some partial information on some variables of the time series vector. Section 4 considers the case of unbinding constraints that includes forecast combination, intercept correction, inclusion of expert judgments and conjectures into the forecasts, and benchmarking. In Section 5 we present some practical considerations about the effect of having used nonlinear transformations of variables and how to deal with the problem of unknown covariance matrices that are required by the combining rule. Section 6 presents an example of combining multivariate forecasts for the components of a time series vector with univariate forecasts for the aggregate of these components and shows that an important increase in precision can be obtained from this combination. Section 7 concludes with some final remarks.

2. A general combining rule

Our main concern is to predict a k-variate random vector of multiple time series $\mathbf{Z}=(\mathbf{Z}'_{T+1},\ldots,\mathbf{Z}'_{T+H})'$ where $H \geqslant 1$ is the forecast horizon. On the one hand we shall assume that a model has been built that provides the forecast vector $\hat{\mathbf{Z}}_X$ based on a set of information X. Usually X includes only past values of the time series vector, but it may also include predetermined values of other exogenous variables, intervention effects and so on. We assume that the forecasts are in general correlated and may have some unknown linear deterministic effects that we call bias. This structure can be written as

$$\hat{\mathbf{Z}}_X = L\mathbf{b} + \mathbf{Z} + \mathbf{e}, \tag{2.1}$$

where L is a $kH \times g$ known matrix, with $g \leqslant kH$, $\mathbf{b}$ is a $g \times 1$ unknown vector of bias parameters and $\mathbf{e}=(\mathbf{e}'_{T+1},\ldots,\mathbf{e}'_{T+H})'$ is a zero-mean random error vector with elements $\mathbf{e}_t=(e_{1t},\ldots,e_{kt})'$ and known covariance matrix $\mathrm{Var}(\mathbf{e}|X)=\Sigma_e$. Thus, we allow for the possibility of bias in the prediction of $\mathbf{Z}$ by means of $\hat{\mathbf{Z}}_X$. In fact, from (2.1) we know that $E(\mathbf{Z}|X)=\hat{\mathbf{Z}}_X - L\mathbf{b}$. If the bias is constant for all $\mathbf{Z}_{T+j}$, for $j=1,\ldots,H$, then $g=k$ and $L=(I_k,\ldots,I_k)$, with I_k the k-dimensional identity matrix. This bias may result from the use of a poor or defective information set X, an incorrect model or errors of measurement. The forecasts are in general correlated and the matrix Σ_e, that depends on the model used, includes both the autocovariances and crosscovariances among the components of the error vector.

On the other hand, we suppose that we also have access to an M-dimensional vector of forecasts, $\mathbf{Y}$, coming from an extra-model source of information. This vector is related to some known linear combinations of the future variables of interests, $C\mathbf{Z}$. Thus, $\mathbf{Y}$ may correspond to:

(1) Forecasts generated by another model with different data frequency. For instance, suppose that the time series $\{\mathbf{Z}_t\}$ corresponds to monthly data and we want to predict $H=12$ months ahead. We also have an aggregate forecast for the vector of time series coming from a multivariate model built from yearly data. Then $\mathbf{Y}=(Y_1,\ldots,Y_k)'$ has the yearly predictions for each component and this additional information corresponds to the sum of the monthly values. In this case $M=k$.

(2) Forecasts coming from another model with different level of aggregation over the variables. For instance, let the elements of the time series vector $\mathbf{Z}_t$ be the components of a price index for which a univariate time series model exists. Calling Y_{T+i} to the univariate forecast for the price index at period $T+i$, the vector $\mathbf{Y}=(Y_{T+1},\ldots,Y_{T+H})'$ provides additional information on the weighted sum of the components for the whole forecasting horizon. Here $M=H$.
(3) More complicated situations in which both types of aggregate information may be available, at least for some horizons and/or for some components. In particular we may have deterministic restrictions that the predictor must fulfill for some particular aggregation scheme over the components and/or over time.

These situations can be modelled by assuming that a vector $\mathbf{Y}=(\mathbf{Y}_1',\ldots,\mathbf{Y}_A')'$ is available, where $\mathbf{Y}_j=(Y_{1j},\ldots,Y_{bj})'$ for $j=1,\ldots,A$, with $b\leqslant k$ and $M=bA$. This vector is related to $\mathbf{Z}$ by way of a set of stochastic (or unbinding) linear restrictions that can be expressed as

$$\mathbf{Y}=C\mathbf{Z}+\mathbf{u}, \tag{2.2}$$

where C is a known $M\times kH$ matrix of rank M, with $M\leqslant kH$, and $\mathbf{u}=(\mathbf{u}_1',\ldots,\mathbf{u}_A')'$ is a random error vector with $\mathbf{u}_j=(u_{1j},\ldots,u_{bj})'$ for $j=1,\ldots,A$, and known covariance matrix $\mathrm{Var}(\mathbf{u}|X)=\Sigma_u$. In general the forecasts will be correlated and we let the two models share some common information by allowing for nonzero crosscorrelations between predictors. However, in what follows we shall assume that $E(\mathbf{u}|X)=\mathbf{0}$, so that the stochastic restrictions implied by (2.2) are unbiased. A more general setting would include $E(\mathbf{u}|X)=\boldsymbol{\delta}\neq\mathbf{0}$ which could be interpreted as bias in the restrictions, but this may lead to an identification problem (see Remark 6 below). Besides, in the problems considered here $\mathbf{Y}$ is usually obtained from a reliable source of information, so that we can safely assume that $\boldsymbol{\delta}=\mathbf{0}$. This is in contrast with the problem of parameter estimation in which the extra-model information (also called outside or prior information) is many times just an analyst's guess that should be tested empirically (see Judge et al., 1980, Chapter 3).

Within this framework we want to obtain the BLUE (Best linear unbiased estimator) of the bias parameter $\mathbf{b}$ and the MMSELP of $\mathbf{Z}$ using all the available information. These are obtained by the following general result which is proved in Appendix A by applying GLS.

General Combining Rule (*GCR*): Let us suppose that the random vectors $\mathbf{Z}$, $\hat{\mathbf{Z}}_X$ and $\mathbf{Y}$ are related by means of (2.1) and (2.2) with the following assumptions: $E(\mathbf{Z})=\mathbf{0}$, $E(\mathbf{e}|X)=\mathbf{0}$, $\mathrm{Var}(\mathbf{e}|X)=\Sigma_e$, $E(\hat{\mathbf{Z}}_X\mathbf{e}'|X)=0$, $E(\mathbf{u}|X)=\mathbf{0}$, $\mathrm{Var}(\mathbf{u}|X)=\Sigma_u$, $E(\hat{\mathbf{Z}}_X\mathbf{u}'|X)=0$ and $E(\mathbf{eu}'|X)=\Sigma_{eu}$. If $\hat{\mathbf{Z}}_X$, $\mathbf{Y}$, Σ_e, Σ_u and Σ_{eu} are known and all the inverse matrices involved exist, then the BLUE of $\mathbf{b},\hat{\mathbf{b}}$, and MMSELP of $\mathbf{Z}$, $E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y})$, as well as their covariance matrices $\Sigma_b=\mathrm{Var}(\hat{\mathbf{b}})$, $\Sigma_{bZ}=\mathrm{Cov}[\hat{\mathbf{b}},E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y})-\mathbf{Z}]$ and $\Sigma_Z=\mathrm{Var}[E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y})-\mathbf{Z}]$, are given by

$$\hat{\mathbf{b}}=-\Sigma_bL'C'\Sigma_d^{-1}(\mathbf{Y}-C\hat{\mathbf{Z}}_X), \tag{2.3}$$

$$E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y})=(\hat{\mathbf{Z}}_X-L\hat{\mathbf{b}})+A_Z[\mathbf{Y}-C(\hat{\mathbf{Z}}_X-L\hat{\mathbf{b}})], \tag{2.4}$$

$$\Sigma_b = (L'C'\Sigma_d^{-1}CL)^{-1}, \tag{2.5}$$

$$\Sigma_{bZ} = -\Sigma_b L'(I_{kH} - A_Z C)' \tag{2.6}$$

and

$$\Sigma_Z = \Sigma_e - A_Z(C\Sigma_e - \Sigma'_{eu}) + (I_{kH} - A_Z C)L\Sigma_b L'(I_{kH} - A_Z C)', \tag{2.7}$$

where

$$\Sigma_d = \mathrm{Var}(\mathbf{u} - C\mathbf{e}|X) = \Sigma_u - C\Sigma_{eu} - \Sigma'_{eu}C' + C\Sigma_e C' \tag{2.8}$$

and

$$A_Z = (\Sigma_e C' - \Sigma_{eu})\Sigma_d^{-1}. \tag{2.9}$$

Remark 1. We assume that, without the historical information X, the vector $\mathbf{Z}$ has expectation zero. This implies no loss of generality because a deterministic known mean can always be subtracted before the analysis. However, given the historical data we have $E(\mathbf{Z}|X) = \hat{\mathbf{Z}}_X - L\mathbf{b}$.

Remark 2. In Appendix A we show that not only is $\hat{\mathbf{b}}$ unbiased for $\mathbf{b}$, but $E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})$ is also an unbiased predictor of $\mathbf{Z}$, in the sense that $E[E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y}) - \mathbf{Z}|X] = \mathbf{0}$, whereas $\hat{\mathbf{Z}}_X$ is biased if $\mathbf{b} \neq 0$.

Remark 3. There are many other ways of expressing (2.3)–(2.9). We chose these particular expressions since they allow us to see the effect of making $\Sigma_u = 0$ directly. A way of interpreting the previous rule is as saying that, before combining $\hat{\mathbf{Z}}_X$ and $\mathbf{Y}$ to forecast $\mathbf{Z}$, we should first correct for bias $\hat{\mathbf{Z}}_X$, using to that end the optimal estimator $\hat{\mathbf{b}}$. Let us also notice that if $\mathbf{Y} = C\hat{\mathbf{Z}}_X$, then $\hat{\mathbf{b}} = \mathbf{0}$ and $E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y}) = \hat{\mathbf{Z}}_X$, so that no additional information is provided by $\mathbf{Y}$ to improve the optimality of $\hat{\mathbf{Z}}_X$ as predictor of $\mathbf{Z}$.

Remark 4. Eq. (2.4) indicates that the optimal forecast is a weighted average of both forecasts, with weights proportional to their precision (see Peña, 1997, for many examples of this rule). To see this, let us assume first that $\Sigma_{eu} = 0$. Then we have that $E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})$ is the sum of $\Sigma_e C'(\Sigma_u + C\Sigma_e C')^{-1}\mathbf{Y}$ and $[\mathbf{I} - \Sigma_e C'(\Sigma_u + C\Sigma_e C')^{-1}C](\hat{\mathbf{Z}}_X - L\hat{\mathbf{b}})$. So that the two unbiased predictors of $C\mathbf{Z}$ and $\mathbf{Z}$, namely $\mathbf{Y}$ and $\hat{\mathbf{Z}}_X - L\hat{\mathbf{b}}$, are weighted according to their precision. When $\Sigma_{eu} \neq 0$ the same rule applies, but now the covariance matrix of each forecast is corrected from the common information contained in Σ_{eu}.

Remark 5. Guerrero and Peña (2000) obtained a particular result of this rule for combining forecasts and linear restrictions in univariate time series, which corresponds to the particular case $k = 1$, $\mathbf{b} = \mathbf{0}$, $\Sigma_{eu} = 0$ and $\Sigma_u = 0$.

Remark 6. To appreciate the convenience of assuming that the restrictions are unbiased, consider a simple univariate situation in which $L=(1,\ldots,1)$ and $C=(0,\ldots,0,1)$. Then (2.1) and (2.2) become $\hat{Z}_{X,T+h}=b+Z_{T+h}+e_{T+h}$ for $h=1,\ldots,H$ and $Y=Z_{T+H}+u$, so that the GCR yields $\hat{b}=\hat{Z}_{X,T+H}-Y$ and $E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y})=\hat{\mathbf{Z}}_X-L\hat{b}$. These results are also obtained by noting that $Y=(\hat{Z}_{X,T+H}-b-e_{T+H})+u$ so that $b=\hat{Z}_{X,T+H}-Y+(u-e_{T+H})$. Thus if we make $u-e_{T+H}=0$ we get $\hat{b}$ as before, in such a way that this plug-in procedure works well in this case. Now let us consider bias in the restriction, i.e. $Y=f+Z_{T+H}+u$. Then the plug-in procedure leads us to $Y=f+(\hat{Z}_{X,T+H}-b-e_{T+H})+u$ and $b-f=\hat{Z}_{X,T+H}-Y+(u-e_{T+H})$ so that $b-f$ can be estimated by $\hat{Z}_{X,T+H}-Y$, but it is impossible to identify the effects of b and f separately.

2.1. Some particularly important simpler combining rules

Let us start by considering a combining rule that comes out from the assumption that $\mathbf{b}=\mathbf{0}$. In fact, most of the examples that we consider in the following sections assume that $\hat{\mathbf{Z}}_X$ is a (conditionally) unbiased forecast for $\mathbf{Z}$. Even if some bias is deemed to be present, the combination procedures employed in practice involve debiasing as a first step and then combining the forecast from different sources. Let us recall that this kind of intuition is justified by the GCR, where the bias parameter vector is jointly estimated with $\mathbf{Z}$. Thus, we now assume in first place that the potential causes of bias have been avoided and that the remaining bias has been removed previously to attempting the linear combination of $\hat{\mathbf{Z}}_X$ and $\mathbf{Y}$. When that happens we get as a result (2.4) and (2.7) to (2.9) with L replaced by the matrix 0. Then such equations can be equivalently expressed as (see Appendix B)

$$E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y})=\Sigma_Z(\Sigma_e^{-1}+\Sigma_f\Sigma_{u|e}^{-1}\Sigma_{eu}'\Sigma_e^{-1})\hat{\mathbf{Z}}_X-\Sigma_Z\Sigma_f\Sigma_{u|e}^{-1}\mathbf{Y} \tag{2.10}$$

with

$$\Sigma_Z=(\Sigma_e^{-1}+\Sigma_f\Sigma_{u|e}^{-1}\Sigma_f')^{-1}, \tag{2.11}$$

where $\Sigma_{u|e}=\Sigma_u-\Sigma_{eu}'\Sigma_e^{-1}\Sigma_{eu}$ and $\Sigma_f=\Sigma_e^{-1}\Sigma_{eu}-C'$.

According to (2.4), to obtain $E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y})$ we should start with the forecast $\hat{\mathbf{Z}}_X$ and adjust it towards $\mathbf{Y}$ by weighting the discrepancy $\mathbf{Y}-C\hat{\mathbf{Z}}_X$ between the two forecasts. We also notice that $CE(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y})=\mathbf{Y}$ when $\Sigma_u=0$ (hence $\Sigma_{eu}=0$ and $\Sigma_{u|e}=0$) or when $C\hat{\mathbf{Z}}_X=\mathbf{Y}$. The former case is particularly interesting because then (2.2) imposes a set of linearly independent restrictions on the prediction of $\mathbf{Z}$. With regard to (2.7), we notice that Σ_e is equal to Σ_Z plus a positive semidefinite matrix. Therefore, the linear combination always improves the efficiency of the estimation when $\Sigma_{eu}=0$. However, when $\Sigma_{eu}\neq 0$, it turns out that Σ_Z can sometimes get larger than Σ_e.

It should also be stressed that in practice, expressions (2.4) and (2.7) are preferable than (2.10) and (2.11) since less matrix inversions are required and the matrices to be inverted are of lower dimension. Moreover, let us notice that when $L=0$, (2.1) yields the (conditionally) unbiased predictor $E(\mathbf{Z}|X)=\hat{\mathbf{Z}}_X$, with precision Σ_e^{-1}, while (2.2) also produces the unbiased predictor $E(\mathbf{Z}|\mathbf{Y})=(C'\Sigma_u^{-1}C)^{-1}C'\Sigma_u^{-1}\mathbf{Y}$ which is correlated with $\hat{\mathbf{Z}}_X$ if $\Sigma_{eu}\neq 0$. However, we can obtain another predictor uncorrelated with $\hat{\mathbf{Z}}_X$

by using the information included in $\mathbf{Y}$ that has no linear relationship with $\hat{\mathbf{Z}}_X$. This additional information is the difference between $\mathbf{Y}$ and $E(\mathbf{Y}|\hat{\mathbf{Z}}_X) = \Sigma'_{eu}\Sigma_e^{-1}\hat{\mathbf{Z}}_X$. Let us call $\hat{\mathbf{Y}} = \mathbf{Y} - E(\mathbf{Y}|\hat{\mathbf{Z}}_X)$ to the residual vector of the regression of $\mathbf{Y}$ on $\hat{\mathbf{Z}}_X$, whose precision is given by $\Sigma_{u|e}^{-1}$. In order to compute the predictor of $\mathbf{Z}$ derived from $\hat{\mathbf{Y}}$ we note that (see Appendix C)

$$E(\mathbf{Z}|X, \hat{\mathbf{Y}}) = -(\Sigma_f \Sigma_{u|e}^{-1}\Sigma'_f)^{-1}\Sigma_f\Sigma_{u|e}^{-1}(\mathbf{Y} - \Sigma'_{eu}\Sigma_e^{-1}\hat{\mathbf{Z}}_X). \tag{2.12}$$

Thus (2.10) can be interpreted as the combination of two unbiased and uncorrelated predictors of $\mathbf{Z}$, namely $E(\mathbf{Z}|X)$ and $E(\mathbf{Z}|X, \hat{\mathbf{Y}})$, with corresponding weighting matrices $\Sigma_Z\Sigma_e^{-1}$ and $\Sigma_Z(\Sigma_f\Sigma_{u|e}^{-1}\Sigma'_f)^{-1}$. Also since

$$\mathrm{Var}[E(\mathbf{Z}|X, \hat{\mathbf{Y}}) - \mathbf{Z}] = (\Sigma_f\Sigma_{u|e}^{-1}\Sigma'_f)^{-1} \tag{2.13}$$

then (2.11) can be interpreted as saying that the precision of $E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})$ is the sum of the precisions associated with the two uncorrelated forecasts. This fact will be exploited below to derive a measure of precision share attributable to each uncorrelated source of information. It is also interesting to realize that Σ_b does not enter the formula for $E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})$ in (2.4), therefore when $\mathbf{b}$ is assumed known, $\hat{\mathbf{Z}}_X - L\mathbf{b}$ in (2.4) plays the role of $\hat{\mathbf{Z}}_X$, as if $L = 0$. However, by assuming $\mathbf{b}$ known, when in fact it is estimated from the data, we would underestimate Σ_b, as can be seen in (2.7) if we incorrectly assume $L = 0$.

A further combining rule that plays an important role in applications is obtained by assuming, in addition to unbiasedness, that the sources of information are uncorrelated. In that situation we have $L = 0$ and $\Sigma_{eu} = 0$. Then

$$E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y}) = \hat{\mathbf{Z}}_X + A_Z(\mathbf{Y} - C\hat{\mathbf{Z}}_X) = \Sigma_Z\Sigma_e^{-1}\hat{\mathbf{Z}}_X + \Sigma_Z C'\Sigma_u^{-1}\mathbf{Y} \tag{2.14}$$

with

$$\Sigma_Z = (I_{kH} - A_Z C)\Sigma_e = (\Sigma_e^{-1} + C'\Sigma_u^{-1}C)^{-1}, \tag{2.15}$$

where A_Z is given by (2.9) with $\Sigma_{eu} = 0$. In this case $E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})$ is given again by a linear combination of two unbiased and uncorrelated predictors, namely $\hat{\mathbf{Z}}_X$ and $(C'\Sigma_u^{-1}C)^{-1}C'\Sigma_u^{-1}\mathbf{Y}$. The corresponding weighting matrices are now $\Sigma_Z\Sigma_e^{-1}$ and $\Sigma_Z(C'\Sigma_u^{-1}C)$, and the precision of $E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})$ is the sum of the precision matrices of those two forecasts. Then, as $\Sigma_u \to 0$ the additional information provided by $\mathbf{Y}$ becomes more precise and the combined estimator gets closer to fulfill exactly the linear restrictions, thus the precision share attributable to $\mathbf{Y}$ should be expected to get larger. Below we derive a measure of precision shares that allow us to see this fact more clearly.

The GCR keeps a strong resemblance with two well-known methods of estimation. The first one, called mixed estimation, was proposed by Theil and Goldberger (1961) (see also Theil, 1974). Those authors were interested in fixed parameter estimation rather than in random vector prediction, but their arguments and derivations are similar to those supporting our combining rule. Besides, the ideas expressed by Theil (1963) led us to compatibility testing and to measuring precision shares. The other closely related method is Bayesian estimation, in which some prior beliefs are expressed through model (2.2) in order to improve both the forecast accuracy and efficiency of $\hat{\mathbf{Z}}_X$. An important feature of Bayesian estimation is that simultaneous inference can be made on both parameters and random variables, for instance on Σ_e and $\mathbf{Z}$, by assuming appropriate prior distributions. Similarly, it is worth mentioning the works of Harvey and his collaborators (e.g. Harvey and Stock, 1993; or Harvey and Chung, 2000) who have advocated the use of structural time series models. Those models together with Kalman filtering can be used to get a solution to the forecast combination problem, that can be recursively implemented.

2.2. Measuring precision shares and testing for compatibility

The idea of measuring precision shares consists of obtaining a scalar index taking values between zero and one, to quantify the precision share of each uncorrelated source of information in the precision of the combined predictor. In order for such a measure to be sensible, we should first ask ourselves whether the observations to be combined are compatible with each other. In particular let us notice that when bias is present, $\hat{\mathbf{Z}}_X$ and $\mathbf{Y}$ provide incompatible data, therefore bias should be avoided or else debiasing should be carried out before combining. When $\mathbf{b}$ can be assumed known and $\Sigma_b = 0$, then $\hat{\mathbf{Z}}_X^* = \hat{\mathbf{Z}}_X - L\mathbf{b}$ plays the role of $\hat{\mathbf{Z}}_X$. Nevertheless, we should be aware that a statistical test of compatibility between $\hat{\mathbf{Z}}_X$ and $\mathbf{Y}$ can be performed as indicated below, thus providing empirical evidence on the adequacy of the combining procedure.

Let us recall that (2.11) shows that the precision of the combined forecast is the sum of the precisions associated with two uncorrelated predictors. Thus, we can apply the original idea of Theil (1963) to obtain measures of precision share attributable to each (uncorrelated) source of information. Those measures are expressed as

$$\mathrm{Prec}[\hat{\mathbf{Z}}_X | E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})] = (kH)^{-1}\,\mathrm{tr}(\Sigma_e^{-1}\Sigma_Z) \tag{2.16}$$

and

$$\mathrm{Prec}[\mathbf{Y} - \Sigma_{eu}'\Sigma_e^{-1}\hat{\mathbf{Z}}_X | \mathbf{E}(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})] = 1 - \mathrm{Prec}[\hat{\mathbf{Z}}_X | \mathbf{E}(\mathbf{Z}|\hat{\mathbf{Z}}_{\mathbf{X}}, \mathbf{Y})]$$
$$= (kH)^{-1}\,\mathrm{tr}[(\Sigma_Z^{-1} - \Sigma_e^{-1})\Sigma_Z], \tag{2.17}$$

where Σ_Z is given either by (2.7) with $L=0$ or by (2.15), depending on whether Σ_{eu} is different from the zero matrix or not. These measures take on values between zero and one, add up to unity and are invariant under nonsingular linear transformations of the variables. Expression (2.16) is to be read as the precision of $E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})$ attributable to $\hat{\mathbf{Z}}_X$ and it can be interpreted as the proportion of MSE of $E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})$ relative to that of $\hat{\mathbf{Z}}_X$. In (2.17) we appreciate that the precision share measures also the relative reduction in MSE attributable to using $\mathbf{Y}$ in addition to $\hat{\mathbf{Z}}_X$. These interpretations are symmetrical with respect to the roles of $\hat{\mathbf{Z}}_X$ and $\mathbf{Y}$ when such vectors are uncorrelated.

We now focus on the implicit assumption behind the idea of combining data from different sources, which is that they provide compatible evidence about the random vector $\mathbf{Z}$. Of course, there may be subject matter considerations pointing in that direction, but in practice it may be a good idea to verify such an assumption empirically. To that end we assume a normal distribution for $\mathbf{e}$ and $\mathbf{u}$ and consider the two conditional means $E(\mathbf{Z}|X)$ and $E(\mathbf{Z}|\mathbf{Y})$ which are fixed once X and $\mathbf{Y}$ are given. Those conditional means are the population characteristics we are interested in. Then we postulate as null hypothesis

$$H_0\colon E(C\mathbf{Z}|X) - E(C\mathbf{Z}|\mathbf{Y}) = \mathbf{0}. \tag{2.18}$$

Let us now recall that $E(C\mathbf{Z}|X) = C\hat{\mathbf{Z}}_X$ and $E(C\mathbf{Z}|\mathbf{Y}) = \mathbf{Y}$ are both assumed to be unbiased for $C\mathbf{Z}$, with $\mathrm{Cov}(C\hat{\mathbf{Z}}_X - C\mathbf{Z}) = C\Sigma_e C'$, $E[(C\hat{\mathbf{Z}}_X - C\mathbf{Z})(\mathbf{Y} - C\mathbf{Z})] = C\Sigma_{eu}$ and $\mathrm{Cov}(\mathbf{Y} - C\mathbf{Z}) = \Sigma_u$. Then, on the null hypothesis, the difference of estimation errors $\mathbf{d} = \mathbf{Y} - C\hat{\mathbf{Z}}_X$, has an M-dimensional normal distribution with $E(\mathbf{d}|X) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{d}|X) = \Sigma_d$ given by (2.8). Hence, we obtain the test statistic

$$K = \mathbf{d}'\Sigma_d^{-1}\mathbf{d} \sim \chi^2_M \tag{2.19}$$

and, given $\hat{\mathbf{Z}}$ and $\mathbf{Y}$, we may calculate the compatibility test statistic

$$K_{\text{calc}} = (\mathbf{Y} - C\hat{\mathbf{Z}}_X)'\Sigma_d^{-1}(\mathbf{Y} - C\hat{\mathbf{Z}}_X). \tag{2.20}$$

A similar statistic was proposed by Guerrero (1989) in a univariate setting and Theil (1974) was led to an equivalent result from a different argument.

Another relevant statistical test may be carried out when bias is allowed to occur. The null hypothesis of unbiasedness is $H_0\colon \mathbf{b} = \mathbf{0}$, for which the following test statistic is derived from the normality of $\hat{\mathbf{b}}$

$$\lambda = \hat{\mathbf{b}}'\Sigma_b^{-1}\hat{\mathbf{b}} \sim \chi^2_g, \tag{2.21}$$

where Σ_b is given by (2.5). Thus, given $\hat{\mathbf{Z}}_X$ and $\mathbf{Y}$, we obtain the calculated statistic

$$\lambda_{\text{calc}} = (\mathbf{Y} - C\hat{\mathbf{Z}}_X)'\Sigma_d^{-1}CL\Sigma_b L'C'\Sigma_d^{-1}(\mathbf{Y} - C\hat{\mathbf{Z}}_X). \tag{2.22}$$

When the null hypothesis is not rejected by the data, we reach the conclusion that both sources of information are compatible with each other in terms of unbiasedness. The alternative hypothesis in this case is just that $\hat{\mathbf{Z}}_X$ is biased. The test based on K is far more general, since its alternative hypothesis is open to such possibilities as: (i) change in the deterministic structure of the model producing $\hat{\mathbf{Z}}_X$ (perhaps due to the presence of constant bias $\mathbf{b}$); (ii) change in the stochastic structure of the corresponding model; (iii) change in the parameter values. These cases have been considered in some detail for univariate time series by Guerrero (1990, 1991). Furthermore, let us notice that $\Sigma_u = 0$ is allowed by both K_{calc} and λ_{calc}. If such were the case, (2.20) would also consider as alternative hypothesis the possibility of having $\Sigma_u \neq 0$, while (2.22) will lead to bias as the only possible explanation.

When either K_{calc} or λ_{calc} leads to rejection of its corresponding null hypothesis, we should also be able to carry out individual or partial tests. This can be done, for $m = 1, \ldots, M$, with the aid of

$$K_{m,\text{calc}} = (\mathbf{Y}_m - C_m\hat{\mathbf{Z}}_X)'(\Sigma_{u,m} - C_m\Sigma_{eu} - \Sigma'_{eu}C'_m + C_m\Sigma_e C'_m)^{-1}(\mathbf{Y}_m - C_m\hat{\mathbf{Z}}_X) \tag{2.23}$$

by selecting $\mathbf{Y}_m$, C_m and $\Sigma_{u,m}$ appropriately. The calculated value should be compared with a Chi-square distribution with k degrees of freedom. In a similar fashion, for each

$i = 1, \ldots, k$, we could use the statistic

$$\lambda_i = \hat{b}_i^2 \Sigma_{b,i}^{-1} \sim \chi_1^2 \tag{2.24}$$

to test H_0: $b_i = 0$, where $\Sigma_{b,i}^{-1}$ is the ith element in the diagonal of Σ_b^{-1}.

3. Linear combination with binding constraints

This section is concerned with estimating a vector of multiple time series when the additional information imposes binding constraints, i.e. when $\Sigma_u = 0$. For univariate time series without bias, that is $k = 1$ and $\mathbf{b} = \mathbf{0}$, this problem was already considered by Guerrero and Peña (2000). In that work we showed that such univariate time series problems as missing data estimation, restricted forecasting with binding constraints, analysis of influential observations and outliers, including reallocation outliers, and temporal disaggregation, can be studied within this common framework. Here we could generalize to the multivariate case the solution to all the univariate problems considered in that paper. As this generalization is straightforward, we only present briefly two extensions not considered previously.

3.1. Deterministic changes in the model structure

We now address the situation in which $\mathbf{b} \neq \mathbf{0}$, indicating essentially that $E(\mathbf{Z}|X)$ and $\mathbf{Y}$ are incompatible. For this case one possibility is to have constant bias in the forecasts provided by $E(\mathbf{Z}|X)$, so that $L = (I_k, \ldots, I_k)$ and $\mathbf{b}$ is $k \times 1$ in model (2.1). A more general approach assumes that

$$L = (\Delta_1, \ldots, \Delta_H) \text{ with } \Delta_h = \mathrm{diag}(\delta_{1h}, \ldots, \delta_{kh}) \text{ for } h = 1, \ldots, H. \tag{3.1}$$

Then the following cases, proposed by Tsay (1988) in a different context, may also be considered appropriate here.

(i) *Pulse effect*. Let $\Delta_h = I_k$ for a particular h value and $\Delta_j = 0$ for $j \neq h$, so that $\delta_{ih} = 1$ and $\delta_{ij} = 0$ for $i = 1, \ldots, k$. In this case $\mathbf{b} = (b_1, \ldots, b_k)'$ measures the effect to occur at the time point $T + h$ that renders $E(\mathbf{Z}|X)$ and $\mathbf{Y}$ incompatible. Notice that h may be specified on the basis of subject matter considerations or from the data themselves, by using the testing procedure suggested below.

(ii) *Level change*. This is the case of constant bias previously considered, for which $\Delta_h = I_k$ (that is $\delta_{ih} = 1$, $i = 1, \ldots, k$) for $h = 1, \ldots, H$.

(iii) *Linear trend*. The specification of L is now given by $\Delta_h = hI_k$, so that $\delta_{ih} = h$ for $h = 1, \ldots, H$. In this case the effect accumulates linearly during the forecast horizon, with slope b_i for $i = 1, \ldots, k$.

(iv) *Transient effect*. We now assume an initial effect b_i that dies out exponentially with known rate of decay $0 < \delta_{i1} < 1$, for $i = 1, \ldots, k$. Then we have $\Delta_h = \Delta_1^{h-1}$ for $h = 1, \ldots, H$, so that $\delta_{ih} = \delta_{i1}^{h-1}$ for $h = 1, \ldots, H$.

(v) *Gradual level change*. This situation corresponds to an increasing effect with initial impact b_i which tends gradually to a new permanent level. Here $\Delta_h = \sum_{j=0}^{h-1} \Gamma^j$ with $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_k)$ where $0 < \gamma_i < 1$, for $i = 1, \ldots, k$, are known. The elements of Δ_h are now $\delta_{ih} = (1-\gamma_i^h)/(1-\gamma_i)$ for $h=1,\ldots,H$ and $b_i/(1-\gamma_i)$ is the new permanent level, for $i=1,\ldots,k$.

A further possibility is to specify the matrix L in such a way that a different case is chosen for each variable of $\mathbf{Z}$. The basic point is that the analyst should be able to postulate a particular form of change in deterministic structure based on subject matter knowledge of the phenomenon under study. In a similar fashion as Tsay (1988) proposed to use likelihood ratio tests to distinguish among different types of changes in univariate time series, we also suggest validating empirically the specification of L. This can be done by using the test statistic (2.22). Thus, for each different specification of L we should calculate its corresponding test statistic. That is, we should calculate λ_{PE} (for pulse effect), λ_{LC} (level change), λ_{LT} (linear trend), λ_{TE} (transient effect) and λ_{GC} (gradual change). The maximum value of those calculated statistics provides empirical evidence that its corresponding effect is the most likely deterministic change that has to occur on the series $\{\mathbf{Z}_t\}$ in order to attain the restriction imposed by $\mathbf{Y}$.

3.2. Partial information on some variables

In practice, data are produced with different time delays, therefore a model including several variables may present a "ragged edge" for forecasting purposes. This means that observations on some variables are already available for the forecast horizon. Wallis (1986) presented a review of the solutions to this problem and derived the optimal forecast based on the normality assumption. Even though he considered the particular case of static forecasts, his solution can be easily generalized to the dynamic forecast situation. In fact he assumed that, in our notation, $\mathbf{Z}_t$ can be partitioned as $\mathbf{Z}_t = (\mathbf{Z}_{1t}', \mathbf{Z}_{2t}')'$ where $\mathbf{Z}_{1t}$ is an M-vector containing the variables for which observations are already available at time $t=T+1$, while the $(k-M)$-vector $\mathbf{Z}_{2t}$ has observations up to time T. Accordingly, the white noise vector of the multivariate time series model, $\{\mathbf{a}_t\}$, and its variance-covariance matrix are partitioned as

$$\mathbf{a}_t = \begin{pmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{pmatrix} \quad \text{and} \quad \Sigma_a = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_2 \end{pmatrix}. \tag{3.2}$$

In this context we can apply the GCR in its form (2.14) and (2.15) with $\Sigma_e = \Sigma_a$, $\Sigma_u = 0$, $\mathbf{Y} = C\mathbf{Z}_{T+1}$ and $C = (I_M, 0)$. Then we obtain

$$\begin{pmatrix} E(\mathbf{Z}_{1,T+1} \mid \hat{\mathbf{Z}}_X, \mathbf{Y}) \\ E(\mathbf{Z}_{2,T+1} \mid \hat{\mathbf{Z}}_X, \mathbf{Y}) \end{pmatrix} = \begin{pmatrix} \mathbf{Y} \\ \hat{\mathbf{Z}}_{X,(2,T+1)} + \Sigma_{12}'\Sigma_1^{-1}(\mathbf{Y} - \hat{\mathbf{Z}}_{X,(1,T+1)}) \end{pmatrix} \tag{3.3}$$

with

$$\Sigma_Z = \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_2 - \Sigma_{12}'\Sigma_1^{-1}\Sigma_{12} \end{pmatrix}. \tag{3.4}$$

These results were derived by Wallis (1986). Other works related to this problem are those of Sandee et al. (1984) and Matthews et al. (1994). In the present situation the compatibility test would serve basically to validate the partial information. Let us also notice that (3.3) and (3.4) are applicable when the partial information is on endogenous variables. Nevertheless, when the partial information is only on exogenous variables, the restricted forecasting technique can be used as well. The distinction being now that only the model that dictates the behavior of the exogenous variables is employed.

4. Use of unbinding constraints

Several important problems involve the use of extra-model information of the form (2.2) with $\Sigma_u \neq 0$. Here we contemplate some of those situations.

4.1. Forecast combination

Combining forecasts can be deemed as an attempt at using the best features of alternative models for the same variables. That explains why forecasts from large econometric models, usually good for medium and long range forecasting, are combined with short-term forecasts from time series models. The problem of finding the best combining weights for the different forecasts to be combined has been approached from many different perspectives. Here we consider the solution obtained by applying the GCR. Thus, on the one hand a time series model produces a forecast of $\mathbf{Z}$ as $E(\mathbf{Z}|X)$ where $X = (\mathbf{Z}_1', \dots, \mathbf{Z}_T')'$ is the historical record of $\{\mathbf{Z}_t\}$. On the other, an alternative model produces a forecast of $C\mathbf{Z}$ given by $\mathbf{Y}$, in such a way that (2.2) applies with $\mathrm{Var}(\mathbf{u}|X) = \Sigma_u$. We might assume that the errors in these models are correlated with each other, so that $\Sigma_{eu} \neq 0$. However, most authors that have written about this topic either have assumed a priori that $\Sigma_{eu} = 0$ (e.g. Greene et al., 1986; Pankratz, 1989) or have advised us against the use of an estimated Σ_{eu}, unless we are sure of its reliability (see Newbold and Granger (1974) or Trabelsi and Hillmer (1989)). In this section we make $\Sigma_{eu} = 0$.

Once Σ_e, Σ_u, $E(\mathbf{Z}|X)$ and $\mathbf{Y}$ are known, we can estimate both the bias vector and $\mathbf{Z} = (\mathbf{Z}_{T+1}', \dots, \mathbf{Z}_{T+H}')'$ via the GCR. Nevertheless, in most applications the model forecasts are assumed to be unbiased, or else debiasing is carried out before combining, so that expressions (2.14) and (2.15) are usually applied. In fact, Doan et al. (1984) proposed an algorithm that essentially computes $\hat{\mathbf{Z}}_X$ as in (2.14), similarly Greene et al. (1986) and Pankratz (1989) derived (2.14) and (2.15) in the multivariate case. Whereas Trabelsi and Hillmer (1989) and Guerrero (1989) also derived them and illustrated their use in the univariate case. It should be noticed that forecasts at different levels of aggregation (say monthly and quarterly) are allowed in the combination.

Let us now consider the situation in which $J \geqslant 1$ alternative models are available besides the original one that produces $\hat{\mathbf{Z}}_X$. These alternative models yield

$$\mathbf{Y}_j = C_j\mathbf{Z} + \mathbf{u}_j \quad \text{for } j = 1, \dots, J \tag{4.1}$$

with C_j a known full rank matrix of dimension $M_j \times kH$. Here we assume that $E(\mathbf{u}_j|X) = \mathbf{0}$, $\mathrm{Var}(\mathbf{u}_j|X) = \Sigma_{u,j}$, $E(\mathbf{u}_i\mathbf{u}_j'|X) = 0$ for $i \neq j$ and $E(\mathbf{u}_j\mathbf{a}'|X) = 0$. Then (2.14) and (2.15) can be extended by Induction to obtain the following combined forecast, given $\mathbf{Y}_1,\ldots,\mathbf{Y}_J$ and $\hat{\mathbf{Z}}_X$

$$E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_J)=\Sigma_{Z,J}\Sigma_e^{-1}\hat{\mathbf{Z}}_X+\Sigma_{Z,J}\sum_{j=1}^{J}C_j'\Sigma_{u,j}^{-1}\mathbf{Y}_j \tag{4.2}$$

with

$$\Sigma_{Z,J}=\left(\Sigma_e^{-1}+\sum_{j=1}^{J}C_j'\Sigma_{u,j}^{-1}C_j\right)^{-1}. \tag{4.3}$$

Furthermore, $E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_J)$ and $\Sigma_{Z,J}$ can be computed recursively for $j=1,\ldots,J$, as follows:

$$\begin{aligned}E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_j)&=E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_{j-1})\\&\quad+A_j[\mathbf{Y}_j-C_jE(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_{j-1})]\end{aligned} \tag{4.4}$$

with $E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_0)=\hat{\mathbf{Z}}_X$ and

$$\Sigma_{Z,j}=(I_{kH}-A_jC_j)\Sigma_{Z,j-1}\quad\text{with } \Sigma_{Z,0}=\Sigma_e, \tag{4.5}$$

where

$$A_j=\Sigma_{Z,j-1}C_j'(C_j\Sigma_{Z,j-1}C_j'+\Sigma_{u,j}^{-1}). \tag{4.6}$$

In this situation, it is interesting to measure the precision due to each of the forecasts involved in the combination. Thus, as in (2.16) we now obtain

$$\text{Prec}[\hat{\mathbf{Z}}_X|E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_J)]=(kH)^{-1}\,\text{tr}(\Sigma_{Z,0}^{-1}\Sigma_{Z,J}) \tag{4.7}$$

and, for $j=1,\ldots,J$.

$$\text{Prec}[\mathbf{Y}_j|E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_J)]=(kH)^{-1}\,\text{tr}[\Sigma_{Z,0}^{-1}(\Sigma_{Z,j-1}-\Sigma_{Z,j})] \tag{4.8}$$

so that the cumulative precision share becomes

$$\begin{aligned}\text{Prec}[\mathbf{Y}_1,\ldots,\mathbf{Y}_j|E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_J)]&=\sum_{i=1}^{j}\text{Prec}[\mathbf{Y}_i|E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_J)]\\&=(kH)^{-1}\,\text{tr}[\Sigma_{Z,0}^{-1}(\Sigma_{Z,0}-\Sigma_{Z,j})].\end{aligned} \tag{4.9}$$

It is also important to test for compatibility between the alternative forecasts. To that end, recursive tests can be carried out by means of a statistic similar in nature to (2.20), that is, for $j=1,\ldots,J$ we get

$$\begin{aligned}K_{\text{calc},j}&=[\mathbf{Y}_j-C_jE(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_{j-1})]'(C_j\Sigma_{z,j-1}C_j'+\Sigma_{u,j})^{-1}\\&\quad\times[\mathbf{Y}_j-C_jE(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}_1,\ldots,\mathbf{Y}_{j-1})],\end{aligned} \tag{4.10}$$

which should be compared with a Chi-square distribution with M_j degrees of freedom.

4.2. Intercept corrections

Forecasting practitioners are used to apply judgmental procedures in order to improve the accuracy of model-based forecasts. One such procedure, advocated in particular by Clements and Hendry (1999), is called intercept correction. It amounts to including a bias-correction term in the model-based forecasts in order to take into account the possibility of a level change occurring in the most recent part of the time series under study or during the forecast horizon. In this case no additional information exists, except for the knowledge that the level change may affect the forecasts.

This situation may lead the analyst to consider the fact that the forecasts obtained from a model for the differenced series vector $\{\mathbf{Z}_t - \mathbf{Z}_{t-1}\}$ are robust against level changes. Thus, a set of forecasts for the differenced series provides the vector of additional information $\mathbf{Y}$ as well as its corresponding variance-covariance matrix of forecast errors Σ_u. These are the inputs required for the GCR in order to robustify the original forecasts $\hat{\mathbf{Z}}_X$ against the presence of constant bias due to a potential level change, where the design matrix is given by $L = (I_k, \ldots, I_k)$ and $\mathbf{b}$ is $k \times 1$. Of course, an extension of this argument leads to consider a model for the twice differenced series $\{\mathbf{Z}_t - 2\mathbf{Z}_{t-1} + \mathbf{Z}_{t-2}\}$ in order to prevent against bias in the forecasts due to a trend change (what Clements and Hendry consider as a growth-rate change). In that case, the design matrix is specified as in Section 3 for a linear trend, when a deterministic change is feared to occur.

It should be stressed that testing for compatibility between the model-based forecasts and the additional information in this context serves to assign statistical significance to the intercept correction. While the estimated bias parameter $\hat{\mathbf{b}}$ yields the required intercept correction, whose significance depends heavily on the type of design matrix employed.

4.3. Combining forecasts with expert judgments or conjectures

In the case of forecast combination we assumed that Σ_u was known, since in fact it could be estimated from the corresponding model that produced $\mathbf{Y}$. We now assume that the vector of additional information comes from expert judgments. In such a case, Σ_u should also be obtained from the same source of information providing $\mathbf{Y}$. For instance, in the empirical example provided by Pankratz (1989), some outside information about one variable in a trivariate time series model was available. Then, a range of error values for the outside data was established from a historical record. An estimate of the standard deviation for those errors was obtained by assuming that the observed range is equivalent to ± 3 standard deviation limits. Similarly, Guerrero and Berumen (1998) were concerned with using outside information in a study aimed at forecasting electricity consumption in a univariate setting. They obtained future consumption expectations from electricity consumers through a survey. Those expectations were originally obtained in qualitative terms and then transformed into quantitative data, from which both $\mathbf{Y}$ and Σ_u were derived.

There are situations in which $\mathbf{Y}$ is only a conjecture to be entertained in a scenario analysis and the analyst is unwilling or unable to specify the elements of Σ_u. In such a case, we can use the statistics (2.20) or (2.23) to select a matrix Σ_u that renders $\mathbf{Y}$ and $\hat{\mathbf{Z}}_X$ compatible (assuming $\mathbf{b}=\mathbf{0}$). For instance, given $\mathbf{Y}_m$, C_m, $\hat{\mathbf{Z}}_{X,m}$ and Σ_e we can select $\Sigma_{u,m}$ in such a way that $K_{m,\text{calc}}$ is just less than the α-percent point of a Chi-square distribution with k degrees of freedom, with α a significance level specified beforehand.

4.4. Benchmarking

A benchmarking situation arises whenever two or more sources of information produce data with different frequencies for the same variable. For instance, we may have yearly and monthly data for the same variable. The more reliable dataset (typically the one observed less frequently) is accepted as benchmark and the other data are adjusted to make them compatible with the benchmark. Benchmarking is widely used by Statistical Agencies, see Denton (1971), Trabelsi and Hillmer (1989), Cholette and Dagum (1994) and Dagum et al. (1998).

To be specific, suppose that a multivariate time series model provides forecasts for the next sn periods of a time series vector and let $\hat{\mathbf{Z}}_X$ be the $snk \times 1$ vector of forecasts. We assume that these forecasts have some unknown bias $\mathbf{b}$ and that the covariance matrix of the one step ahead forecast is known from the estimated model. This information is summarized by (2.1). There is also another multivariate (perhaps a simultaneous equation) model that provides n forecasts for the sum over s periods of the same time series vector. For instance, if the first model uses monthly data and the second yearly data, then $s=12$. Let the nk-vector of forecasts be $\mathbf{Y}$, so that (2.2) applies with C equal to the required matrix that sums the monthly data to obtain the yearly figures, and Σ_u is obtained from the yearly model. Then we immediately obtain a solution to the benchmarking problem by applying Eqs. (2.3)–(2.9). Some remarks are now in order.

Remark 7. Suppose that the forecast errors from both models are uncorrelated, i.e. $\Sigma_{eu}=0$, and assume that the bias is constant for each time series component, so that $\mathbf{b}=(b_1,\ldots,b_k)'$. If we apply the GCR, in this case with $k=1$, we obtain the method proposed by Cholette and Dagum (1994) for univariate times series The generalization for a vector of time series is given by the equations

$$\hat{\mathbf{b}} = -[L'C'(\Sigma_u + C\Sigma_e C')^{-1}CL]^{-1}L'C'(\Sigma_u + C\Sigma_e C')^{-1}(\mathbf{Y} - C\hat{\mathbf{Z}}_X) \tag{4.11}$$

and

$$E(\mathbf{Z}|\hat{\mathbf{Z}}_X,\mathbf{Y}) = (\hat{\mathbf{Z}}_X - L\hat{\mathbf{b}}) + \Sigma_e C'(\Sigma_u + C\Sigma_e C')^{-1}[\mathbf{Y} - C(\hat{\mathbf{Z}}_X - L\hat{\mathbf{b}})]. \tag{4.12}$$

Although these two equations seem complicated, the resulting procedure is in fact usually very simple. Suppose for instance that we have a trivariate vector of quarterly time series and that the model that we have built produces biased forecasts. We also have a model built with yearly data that produces unbiased predictions. We generate eight-quarter ahead forecasts with the first model and benchmark them with two yearly predictions generated with the second model. In this case $k=3$, $s=4$, $n=2$ and the matrices C and L are

$$C = \begin{pmatrix} I_3 & I_3 & I_3 & I_3 & 0_3 & 0_3 & 0_3 & 0_3 \\ 0_3 & 0_3 & 0_3 & 0_3 & I_3 & I_3 & I_3 & I_3 \end{pmatrix},$$

$$L' = (\, I_3 \; I_3 \; I_3 \; I_3 \; I_3 \; I_3 \; I_3 \; I_3 \,), \tag{4.13}$$

where 0_3 is the square three-dimensional zero matrix. Then it is straightforward to check that

$$\hat{\mathbf{b}} = -\frac{1}{4}\left(\sum_{i=1}^{2}\sum_{j=1}^{2} P_{ij}\right)^{-1} \begin{bmatrix} \left(\sum_{i=1}^{2} P_{i1}\right)\mathbf{d}_1 \\ \left(\sum_{i=1}^{2} P_{i2}\right)\mathbf{d}_2 \end{bmatrix}, \tag{4.14}$$

where P_{ij} for $i,j = 1,2$ are the 3×3 squared sub-matrices of the precision matrix $(\Sigma_u + C\Sigma_e C')^{-1}$ and $\mathbf{d}_1$ ($\mathbf{d}_2$) is the three-dimensional vector of differences between the yearly forecast of each component for the first (second) year and the sum of the four quarterly forecasts of this component for the same year. This difference is first split evenly among the four quarters and then distributed according to the relative precision of each forecast.

Remark 8. Suppose that both forecasts are unbiased. Then the covariance matrix of the corrected forecast is given by

$$\Sigma_Z = (\Sigma_e^{-1} + C'\Sigma_u^{-1}C)^{-1} \tag{4.15}$$

showing that benchmarking increases the precision of the original forecast by adding to the precision of the most frequent forecast, the precision of the benchmarking forecast.

5. Practical considerations

In this section we consider some situations that may arise in practice and suggest some ways to deal with them.

5.1. Nonlinear transformation of variables

It sometimes happens that the linear expression (2.1) relating the original forecast to the true series holds only for a transformation of the variables. Typically the logarithmic transformation is applied to stabilize the variance of some elements of $\mathbf{Z}$. Other transformations employed in practice include the power or Box-Cox transformations as well as products or ratios of variables. In this situation a multiple time series model produce forecasts for the vector of transformed variables, $T(\mathbf{Z})$. Then (2.1) becomes

$$\widehat{T(\mathbf{Z})}_X = L\mathbf{b} + T(\mathbf{Z}) + \mathbf{e}, \tag{5.1}$$

where $\widehat{T(\mathbf{Z})}_X$ denotes the forecast vector based on X, so that $E[T(\mathbf{Z})_X] = \widehat{T(\mathbf{Z})}_X - L\mathbf{b}$. Besides, $\mathbf{e}$ satisfies all the assumptions in the statement of the GCR.

Usually the available restrictions apply to the original variables, in such a way that (2.2) holds true. Then the easiest solution to the problem of incorporating the linear restrictions into the nonlinear forecasts is to write the restrictions in terms of the transformed variables, whenever that is possible. That is, write

$$\mathbf{Y}^* = C^* T(\mathbf{Z}) + \mathbf{u}^* \tag{5.2}$$

and apply formulas (2.3)–(2.9) in the transformed scale of the variables involved. This solution can be employed only in a few cases, for instance when $C = I$, in which case (5.2) imposes isolated restrictions on the elements of $T(\mathbf{Z})$. Then we have $\mathbf{Y}^* \doteq T(\mathbf{Y})$ and $\mathbf{u}^* = f(\mathbf{u}, \mathbf{Z})$. Finally bring the MMSELP $E[T(\mathbf{Z})|\widehat{T(\mathbf{Z})}_X, \mathbf{Y}]$ to the original scale by applying the inverse transformation, i.e. by calculating $T^{-1}\{E[T(\mathbf{Z})|\widehat{T(\mathbf{Z})}_X, \mathbf{Y}]\}$ on the assumption that the inverse transformation exists. When every element of $T(\mathbf{Z})$ is a power transformation that yields an approximate symmetric distribution for each variable in $\mathbf{Z}$, this predictor can be interpreted as a vector of medians of the marginal distributions of $\mathbf{Z}|\widehat{T(\mathbf{Z})}_X, \mathbf{Y}$. This happens because in that case $E[T(\mathbf{Z})|\widehat{T(\mathbf{Z})}_X, \mathbf{Y}]$ is also the vector of medians of the marginal distributions of $T(\mathbf{Z})|\widehat{T(\mathbf{Z})}_X, \mathbf{Y}$.

A more general approach will consider a first-order approximation of $T(\mathbf{Z})$ by expanding it around $\widehat{\mathbf{Z}}_X = T^{-1}[\widehat{T(\mathbf{Z})}_X]$, which is not necessarily an optimal forecast. Thus we have

$$T(\mathbf{Z}) \doteq T(\hat{\mathbf{Z}}_X) + J(\hat{\mathbf{Z}}_X)(\mathbf{Z} - \hat{\mathbf{Z}}_X) \tag{5.3}$$

with $J(\hat{\mathbf{Z}}_X)$ being the Jacobian of the transformation, evaluated at $\mathbf{Z} = \hat{\mathbf{Z}}_X$. Then, by definition of $\hat{\mathbf{Z}}_X$ we get

$$\hat{\mathbf{Z}}_X - \mathbf{Z} \doteq J^{-1}(\hat{\mathbf{Z}}_X)[\widehat{T(\mathbf{Z})}_X - T(\mathbf{Z})] = J^{-1}(\hat{\mathbf{Z}}_X)L\mathbf{b} + J^{-1}(\hat{\mathbf{Z}}_X)\mathbf{e}, \tag{5.4}$$

where the last equality follows from (5.1). Now let us write this expression as

$$\hat{\mathbf{Z}}_X \doteq L^*\mathbf{b} + \mathbf{Z} + \mathbf{e}^* \tag{5.5}$$

with $L^* = J^{-1}(\hat{\mathbf{Z}}_X)L$ and $\mathbf{e}^* = J^{-1}(\hat{\mathbf{Z}}_X)\mathbf{e}$, so that

$$\mathrm{Var}(\mathbf{e}^*|X) = J^{-1}(\hat{\mathbf{Z}}_X)\Sigma_e J^{-1}(\hat{\mathbf{Z}}_X)' = \Sigma_e^* \tag{5.6}$$

and

$$E(\mathbf{e}^*\mathbf{u}|X) = J^{-1}(\hat{\mathbf{Z}}_X)\Sigma_{eu} = \Sigma_{eu}^*. \tag{5.7}$$

Then we are in such a position that the GCR applies to (5.5) and (2.2) yielding the desired approximate estimator $\hat{\mathbf{b}}$ and predictor $E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})$, with L^*, Σ_e^* and Σ_{eu}^* in place of L, Σ_e and Σ_{eu}. Of course, if a predictor is needed in the transformed scale we can use

$$E[T(\mathbf{Z})|\widehat{T(\mathbf{Z})}_X, \mathbf{Y}] \doteq T[E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y})] \quad \text{and} \quad \Sigma_{T(Z)} = J(\hat{\mathbf{Z}}_X)\Sigma_Z J(\hat{\mathbf{Z}}_X)'. \tag{5.8}$$

5.2. Unknown covariance matrices

The GCR can be applied once Σ_e, Σ_u and Σ_{eu} are either known on a priori grounds or estimated from the available data. If they were known we would be using indeed GLS, otherwise we would be forced to employ a feasible procedure, and most likely we would use estimated generalized least squares (EGLS). The asymptotic properties of EGLS can be derived as in Judge et al. (1980) and, for finite samples, via Monte Carlo simulation, although these results will be highly dependent on specific sample size and model considerations. Here we will not pursue this topic any further.

We now concentrate on the problem of estimating the required covariance matrices involved in the specialized cases of the GCR. In first place, the error covariance matrix Σ_e will always be needed, even in the case of binding constraints where $\Sigma_u = 0$ and $\Sigma_{eu} = 0$. To see how that matrix can be estimated, let us assume that the multiple time series model admits the following finite vector auto-regressive (VAR) representation

$$\mathbf{Z}_t = \mathbf{D}_t + \Pi_1 \mathbf{Z}_{t-1} + \cdots + \Pi_p \mathbf{Z}_{t-p} + \mathbf{a}_t, \tag{5.9}$$

where $\mathbf{D}_t$ is a vector of predetermined variables in the information set X and $\Pi_1, \ldots, \Pi_p$ are matrices of constant parameters. The error $\mathbf{a}_t$ is assumed to follow a Gaussian white noise process with $E(\mathbf{a}_t|X) = \mathbf{0}$ and $E(\mathbf{a}_t\mathbf{a}_t'|X) = \Omega$, then the forecasts produced by this model for $h = 1, \ldots, H$, with origin at $t = T$, are

$$\hat{\mathbf{Z}}_{X,T+h} = \mathbf{D}_{T+h} + \Pi_1 \hat{\mathbf{Z}}_{X,T+h-1} + \cdots + \Pi_p \hat{\mathbf{Z}}_{X,T+h-p} \tag{5.10}$$

with $\hat{\mathbf{Z}}_{X,T+i} = \mathbf{Z}_{T+i}$ for $i \leqslant 0$. If we write $\mathbf{a} = (\mathbf{a}_{T+1}, \ldots, \mathbf{a}_{T+H})'$ the vector of forecast errors becomes

$$\hat{\mathbf{Z}}_X - \mathbf{Z} = \Psi \mathbf{a}, \tag{5.11}$$

where Ψ is a $kH \times kH$ lower triangular matrix with $-I_k$ in the main diagonal, $-\Pi_1$ in the first subdiagonal, $-\Pi_2$ in the second subdiagonal and so on. From (5.11) it follows that $E(\hat{\mathbf{Z}}_X - \mathbf{Z}|X) = 0$ and $\mathrm{Var}(\hat{\mathbf{Z}}_X - \mathbf{Z}|X) = \Psi\Sigma_a\Psi'$. Since in practice we are usually unsure whether (5.9) is an appropriate model or that we are using an appropriate information set, we assume here that $\hat{\mathbf{Z}}_X - \mathbf{Z} = L\mathbf{b} + \mathbf{e}$, with $\mathbf{b}$ unknown and nonestimable if only X is used. Thus we are now in the situation originally considered by expression (2.1), so that $\Sigma_e = \Psi\Sigma_a\Psi'$. Therefore estimating Σ_e becomes an easy task because both Ψ and Σ_a can be estimated by standard methods (see Lütkepohl, 1991) from the available data.

The second important case occurs when the restrictions are unbinding and uncorrelated with the vector $\mathbf{e}$, that is, $\Sigma_u \neq 0$ and $\Sigma_{eu} = 0$. This situation may happen when the two sources of information providing $\hat{\mathbf{Z}}_X$ and $\mathbf{Y}$ are independent, for instance when the restrictions are basically conjectures or expert opinions about some linear combinations of the future values. Then the same source of information that provides the linear restrictions should be exploited to get an estimate of Σ_u. For example, in Pankratz (1989) a forecast manager provided an expert opinion of some sales variable on the basis of sales force information. Moreover, the manager's experience led him/her to the assertion that "the maximum error associated with fields estimates of this type has been $\pm 10\%$". Thus, on the assumption that the expected error range is covered by ± 3 standard deviations, the manager's input was used to estimate Σ_u.

Another situation in which zero correlation between $\mathbf{e}$ and $\mathbf{u}$ can be reasonably and conveniently assumed is when combining forecasts from different models. In fact, as indicated in Section 4.1, the literature on this topic recommends making $\Sigma_{eu} = 0$ (see Newbold and Granger (1974) or Trabelsi and Hillmer (1989)). Nevertheless, the estimation of Σ_u is an integral part when producing forecasts $\mathbf{Y}$ with a statistical model different from (5.9), so that $\hat{\Sigma}_u$ can be obtained again from the same source of information that provides $\mathbf{Y}$. Similarly when benchmarking time series it is customary to assume that $\Sigma_{eu} = 0$, while Σ_u is estimated from the same survey data that produced $\mathbf{Y}$, as indicated by Cholette and Dagum (1994). The next section shows how this matrix can be computed in practice, although we have not found in the example any advantage by assuming that this matrix is different from zero.

6. An empirical example

As an example of the previous rules we will show how to combine the multiple time series forecasts of the components of a vector time series and the forecasts for the aggregate of these components. We consider a vector of five time series, $\mathbf{Z}_t = (Z_{1t}, Z_{2t}, Z_{3t}, Z_{4t}, Z_{5t})'$, in which the components are the cost of living indexes for food, Z_{it}, manufacturing products, Z_{2t}, services, Z_{3t}, hotels and tourism, Z_{4t}, and housing, Z_{5t}, in Spain. The sample data include 309 monthly observations from 1/1976 to 9/2001 of these components and can be obtained from www.ine.es. The first 273 observations (1/1976 to 9/1998) have been used to fit the models and the other 36 are used to evaluate the forecasts. These five components are shown in Fig. 1.

The vector of time series is fitted by a seasonal VARIMA $(2,1,0)(0,1,1)_{12}$

$$(I - \Phi_1 B - \Phi_2 B^2)(I - B^{12})(I - B)\mathbf{Z}_t = (I - \Theta_{12} B^{12})\boldsymbol{\varepsilon}_t, \tag{6.1}$$

where B is the backshift operator such that $BZ_t = Z_{t-1}$ for every Z and t, and $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$. The maximum likelihood estimates are

$$\hat{\Phi}_1 = \begin{pmatrix} 0.382 & 0 & 0 & 0 & 0 \\ 0.182 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.168 & 0 & 0.129 \\ 0 & 0 & 0.214 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.384 \end{pmatrix}, \quad \hat{\Phi}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0.178 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0.135 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.159 \end{pmatrix},$$

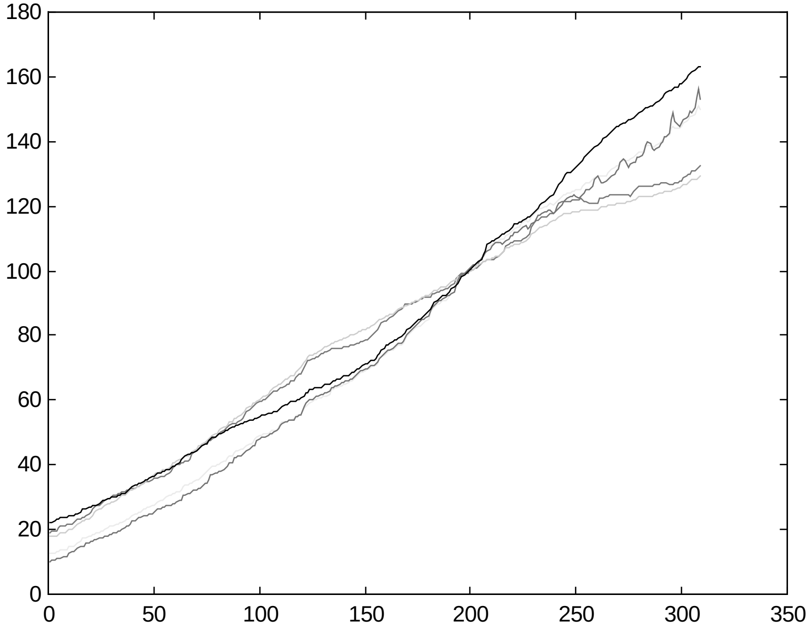

Fig. 1. Five components of the cost of living index in Spain (1976–1998).

$$\hat{\Theta}_{12} = \begin{pmatrix} 0.994 & 0 & -0.095 & 0 & 0.237 \\ 0 & 0.860 & 0 & 0 & 0.139 \\ 0.219 & 0 & 0.712 & 0 & 0 \\ 0.217 & 0 & 0 & 0.666 & 0 \\ 0 & 0.155 & 0 & 0 & 0.802 \end{pmatrix}$$

and

$$\hat{\Sigma}_\varepsilon = \begin{pmatrix} 0.0700 & 0.0286 & -0.0038 & 0.0009 & 0.0001 \\ 0.0286 & 0.0236 & 0.0010 & 0.0084 & 0.0009 \\ -0.0038 & 0.0010 & 0.0555 & 0.0609 & 0.0022 \\ 0.0009 & 0.0084 & 0.0609 & 0.1249 & 0.0023 \\ 0.0001 & 0.0009 & 0.0022 & 0.0023 & 0.0281 \end{pmatrix}. \qquad (6.2)$$

The large value of the moving average parameter at position $(1,1)$ suggests that the first component may be seasonally overdifferenced (the largest real eigenvalue of the $\hat{\Theta}_{12}$ matrix is 0.98). We have also fitted a model in which the first component does not have a seasonal difference but the results were very similar, so we decided to present only the results for the simplest model. The matrix $\hat{\Sigma}_\varepsilon$ is well conditioned (its condition number is 4.14) and there is no strong evidence of common factors or cointegration in the data.

We now form the aggregate univariate time series $Y_t = \sum_{i=1}^{5} Z_{it}$. We decided to use the arithmetic mean of the components instead of a weighting function of them (as in the cost of living index) in order to simplify the presentation and to have a neutral standard to judge the gain in precision when incorporating the aggregate. The same analysis can be applied to any weighting function of the components. The 273 data observations of the aggregated univariate time series have been fitted by the seasonal ARIMA $(2,1,0)(0,1,1)_{12}$ model

$$(1 - 0.225B - 0.191B^2)(1 - B^{12})(1 - B)Y_t = (1 - 0.664B^{12})\hat{a}_t, \tag{6.3}$$

where $\hat{\sigma}_a^2 = 0.6927$. Both models are broadly in agreement with each other because as $Y_t = \mathbf{1}'\mathbf{Z}_t$, then $(1-B^{12})(1-B)Y_t = \mathbf{1}'(I-B^{12})(I-B)\mathbf{Z}$, with $\mathbf{1}$ the 5-dimensional vector of ones. The second term is the sum of MA(∞) components, obtained by inverting the AR(2) matrix operator, leading to a univariate MA(∞) for Y_t, that is in agreement with the ARMA(2,1) univariate model. Also $\hat{\sigma}_a^2 = 0.6927$ is much larger than trace $(\hat{\Sigma}_\varepsilon) = 0.3021$, as expected, due to the presence of several relatively large off-diagonal elements in the parameter matrices of the multivariate model. This agrees with the result of Wei and Abraham (1981) that forecasting the components and adding them to forecast the aggregate dominates in MSE the forecasts for the aggregate obtained with the univariate model (see also Pino et al., 1987).

We generated 3 years of forecasts with the multivariate and the univariate models. Let $\mathbf{Z} = (\mathbf{Z}'_{T+1}, \ldots, \mathbf{Z}'_{T+36})'$ be the vector of dimension 180×1 to be forecasted from the multivariate model and $\hat{\mathbf{Z}}_X = (\hat{\mathbf{Z}}'_{T+1}, \ldots, \hat{\mathbf{Z}}'_{T+36})'$ be the corresponding vector of forecasts based on $X = (\mathbf{Z}_1, \ldots, \mathbf{Z}_{273})$. Also let $\mathbf{Y} = (\hat{Y}_{T+1}, \ldots, \hat{Y}_{T+36})'$ be the vector of univariate forecasts. Computing the combined forecasts requires the covariance matrices of the forecast errors from both models. The covariance matrix $\hat{\Sigma}_e$ of the vector $\mathbf{e} = \hat{\mathbf{Z}}_X - \mathbf{Z}$ of multivariate forecast errors is built from the following squared $k \times k$ block matrices: (1) $\widehat{\mathrm{Var}}(\mathbf{e}_{T+l})$, which includes the variances and covariances of the forecast errors in the components of the time series vector at time $T + l$ and (2) $\widehat{\mathrm{Cov}}(\mathbf{e}_{T+l}, \mathbf{e}_{T+l+i})$, which includes the covariances between the forecast errors at times $T + l$ and $T + l + i$. In order to show the expression for these matrices, let us write the multivariate model as $\mathbf{Z}_t = \hat{\Psi}(B)\hat{\varepsilon}_t$, where the $\hat{\Psi}_j$ matrices are obtained from $(I - \hat{\Phi}_1 B - \hat{\Phi}_2 B^2)(I - B^{12})(I - B) = \hat{\Psi}(B)(I - \hat{\Theta}_{12}B^{12})$. Then $\hat{\mathbf{Z}}_{T+l} = \sum_{j=l}^{\infty} \hat{\Psi}_j \hat{\varepsilon}_{t+l-j}$ and the forecast error for $\mathbf{Z}_{T+l}$ will be given by $\mathbf{e}_{T+l} = \mathbf{Z}_{T+l} - \hat{\mathbf{Z}}_{T+l} = \sum_{j=0}^{l-1} \hat{\Psi} \hat{\varepsilon}_{t+l-j}$, so that $\widehat{\mathrm{Var}}(\mathbf{e}_{T+l}) = \widehat{\mathrm{Var}}(\sum_{j=0}^{l-1} \hat{\Psi}_j \hat{\varepsilon}_{t+l-j}) = \sum_{j=0}^{l-1} \hat{\Psi}_j \hat{\Sigma}_\varepsilon \hat{\Psi}'_j$. Also,

$$\widehat{\mathrm{Cov}}(\mathbf{e}_{T+l}, \mathbf{e}_{T+l+i}) = \widehat{\mathrm{Cov}}\left(\sum_{j=0}^{l-1} \hat{\Psi}_j \hat{\varepsilon}_{T+l-j}, \sum_{j=0}^{l+i-1} \hat{\Psi}_j \hat{\varepsilon}_{T+l+i-j}\right) = \sum_{j=0}^{l-1} \hat{\Psi}_j \hat{\Sigma}_\varepsilon \hat{\Psi}'_{j+i}. \tag{6.4}$$

The computation of the elements of $\hat{\Sigma}_u$, the forecast error covariance matrix of $\mathbf{u} = \mathbf{Y} - C\mathbf{Z}$, requires obtaining $\widehat{\mathrm{Var}}(u_{T+l})$ and $\widehat{\mathrm{Cov}}(u_{T+l}, u_{T+l+i})$, and this is done as follows. Let us write

$$\mathbf{Y} = \begin{pmatrix} \hat{Y}_{T+1} \\ \vdots \\ \hat{Y}_{T+36} \end{pmatrix} = \begin{pmatrix} 1,\ldots,1 & \cdots & 0,\ldots,0 \\ \vdots & \ddots & \vdots \\ 0,\ldots,0 & \cdots & 1,\ldots,1 \end{pmatrix} \begin{pmatrix} \mathbf{Z}_{T+1} \\ \vdots \\ \mathbf{Z}_{T+36} \end{pmatrix} + \mathbf{u} \tag{6.5}$$

and $\hat{Y}_{T+l} = E[Y_{T+l}|Y_T, Y_{T-1},\ldots] = \sum_{j=l}^{\infty} \hat{\psi}_j \hat{a}_{t+l-j}$, where $Y_t = \hat{\psi}(B)\hat{a}_t$, and $\hat{\psi}(B)$ is obtained by equating coefficients of powers of B in $(1-0.225B-0.191B^2)\,(1-B^{12})(1-B) = \hat{\psi}(B)(1-0.664B^{12})$. The univariate forecast error is given by $u_{T+l} = Y_{T+l} - \hat{Y}_{T+l} = \sum_{j=0}^{l-1} \hat{\psi}_j \hat{a}_{t+l-j}$ and has $\widehat{\mathrm{Var}}(u_{T+l}) = \hat{\sigma}_a^2 \sum_{j=0}^{l-1} \hat{\psi}_j^2$ and $\widehat{\mathrm{Cov}}(u_{T+l}, u_{T+l+i}) = \hat{\sigma}_a^2 \sum_{j=0}^{l-1} \hat{\psi}_j \hat{\psi}_{j+i}$. The matrix Σ_{eu} can be estimated by

$$\begin{aligned} \hat{\Sigma}_{eu}(i, i+h) &= \hat{E}(\mathbf{e}_{T+i} u_{T+i+h}) \\ &= \widehat{\mathrm{Cov}}[(\hat{\boldsymbol{\varepsilon}}_{T+i} + \hat{\Psi}_1 \hat{\boldsymbol{\varepsilon}}_{T+i-1} + \cdots + \hat{\Psi}_{i-1} \hat{\boldsymbol{\varepsilon}}_{T+1}) \\ &\quad \times (\hat{a}_{T+i+h} + \hat{\psi}_1 \hat{a}_{T+i+h-1} + \cdots + \hat{\psi}_{i+h-1} \hat{a}_{T+1})] \\ &= \sum_{s=0}^{i-1} \hat{\Psi}_s \left(\sum_{r=0}^{i+h-1} \hat{\psi}_r \, \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\varepsilon}}_{T+i-s}, \hat{a}_{T+i+h-r}) \right), \end{aligned} \tag{6.6}$$

but the covariances were small and we did not find any advantage from using them. Thus we only provide here the results that came out by assuming that $\Sigma_{eu} = 0$.

Using the precision matrices from both forecasts we calculated the Chi-squared compatibility test statistic given by (2.20). The value in this case is 6.89, showing compatibility between both sets of forecasts. The precision share of the multivariate versus the univariate forecasts given by (2.16) is equal to 0.9104. So we may conclude that most of the precision comes from the joint information in the components, although there is some information in the aggregate that can be used to improve the forecasts of the components. A measure of the advantage of incorporating the univariate forecasts to the multivariate ones, in the forecast of each component, can be computed as the relative reduction in squared forecast error for each component. Calling e_{Mjl} to the forecast error of the jth component at time $T+l$ obtained by using the multivariate model, and e_{Cjl} to the forecast error of the same component at time $T+l$, produced by the combination of forecasts, we have that the relative reduction (or increase) of forecast error is given by

$$g(j) = \frac{\sum_{l=1}^{H} e_{Mjl}^2 - \sum_{l=1}^{H} e_{Cjl}^2}{\max(\sum_{l=1}^{H} e_{Mjl}^2, \sum_{l=1}^{H} e_{Cjl}^2)} \tag{6.7}$$

thus a positive (negative) value of $g(j)$ indicates a relative reduction (increase) in the forecast error of the component by using the combined forecast. The global gain is

Table 1
Percentage of improvement in MSE for each component and overall

$g(1)$	$g(2)$	$g(3)$	$g(4)$	$g(5)$	g_T	g_m
15.80	30.10	34.10	29.07	− 12.46	12.38	19.32

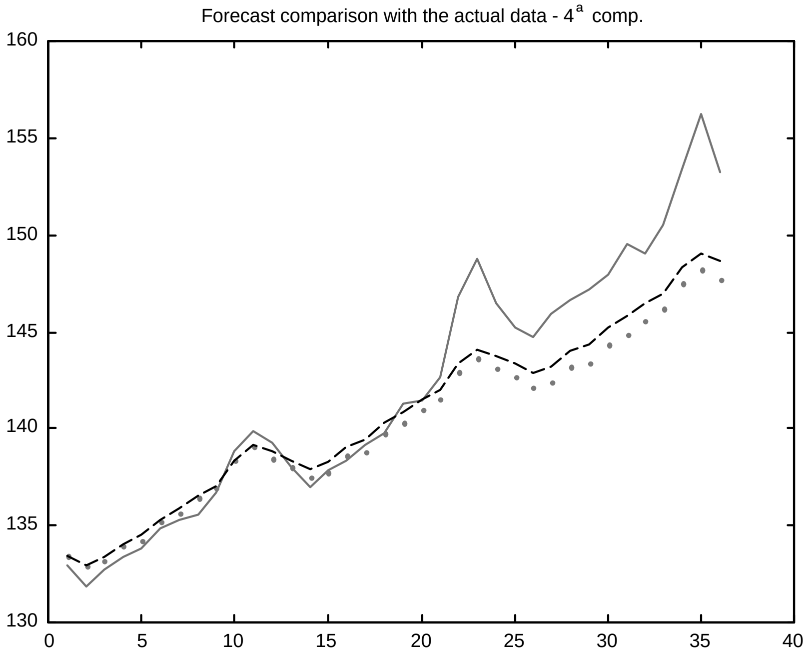

Fig. 2. Observed data (————), Multivariate forecast (. . . .) and combined forecast (— — —) for the 4th component of the time series.

obtained by adding the gains in the five components as follows:

$$g_T = \frac{\sum_{j=1}^{5}\sum_{i=1}^{H} e_{Mji}^2 - \sum_{j=1}^{5}\sum_{i=1}^{H} e_{Cji}^2}{\max(\sum_{j=1}^{5}\sum_{i=1}^{H} e_{Mji}^2 - \sum_{j=1}^{5}\sum_{i=1}^{H} e_{Cji}^2)}. \tag{6.8}$$

An alternative measure of global gain is $g_m = \sum_{j=1}^{5} g(j)/5$. Of course other measures can be used, see for instance Clements and Hendry (1993). The values for $g(j)$, g_T and g_m are indicated in Table 1. The improvements in forecast precision for components 2, 3 and 4 are quite large, and the global gain of incorporating the univariate forecasts is between 12% and 19%, depending on the measure used. Fig. 2 shows a plot of the multivariate and combined forecasts for the 4th component (the median of the values for the components, see Table 1). It can be seen that gain in precision by incorporating the joint information provides better forecasts of the trend in that component.

7. Concluding remarks

We propose the GCR as a unifying tool to solve many apparently unrelated problems. Our approach is theoretically supported by generalized least squares and it was shown to be flexible enough to accommodate problems of estimation and testing within a unified framework. We do not attempt to conclude that the optimal solutions deduced from the GCR are efficient in computational terms. To that end we could employ a recursive approach like Kalman filtering (see, for instance, Harvey and Chung, 2000) known to be computationally more efficient. However, we do believe that the new interpretation of the results previously appeared in the literature is enlightening of such issues as the role of bias, compatibility testing and measurement of precision shares.

It is important to realize that the GCR may be applied to a very general class of multiple time series models that can be put into a reduced form. On the other hand, it should be stressed that many other problems, different from those considered explicitly in Sections 3 and 4, can also be posed in the context of the GCR. Therefore, its limits of application are not just those shown here. In fact, it is up to the analyst's ingenuity to discover more practical applications.

Acknowledgements

The first author's participation was made possible through a professorship in time series analysis and forecasting in econometrics provided by Asociación Mexicana de Cultura, A.C. The second author acknowledges support from DGYCIT grant PB96-0111 and from the Cátedra BBVA de Calidad, Universidad Carlos III de Madrid. We are very grateful to Pedro Galeano who provided great help in the computations for the empirical example.

Appendix A. Proof of the general combining rule

Eqs. (2.1) and (2.2) are put together to form the system

$$\begin{pmatrix} \hat{\mathbf{Z}}_X \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} L & I \\ 0 & C \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{Z} \end{pmatrix} + \begin{pmatrix} \mathbf{e} \\ \mathbf{u} \end{pmatrix} \tag{A.1}$$

with

$$E\begin{pmatrix} \mathbf{e} \\ \mathbf{u} \end{pmatrix} |X) = \mathbf{0} \quad \text{and} \quad \mathrm{Var}\begin{pmatrix} \mathbf{e} \\ \mathbf{u} \end{pmatrix} |X) = \begin{pmatrix} \Sigma_e & \Sigma_{eu} \\ \Sigma'_{eu} & \Sigma_u \end{pmatrix}. \tag{A.2}$$

Therefore, GLS produces

$$\begin{pmatrix} \hat{\mathbf{b}} \\ E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y}) \end{pmatrix} = \begin{pmatrix} \Sigma_b & \Sigma_{bZ} \\ \Sigma'_{bZ} & \Sigma_Z \end{pmatrix} \begin{pmatrix} L' & 0 \\ I & C' \end{pmatrix} \begin{pmatrix} \Sigma_e & \Sigma_{eu} \\ \Sigma'_{eu} & \Sigma_u \end{pmatrix}^{-1} \begin{pmatrix} \hat{\mathbf{Z}}_X \\ \mathbf{Y} \end{pmatrix} \quad (A.3)$$

with

$$\begin{pmatrix} \Sigma_b & \Sigma_{bZ} \\ \Sigma'_{bZ} & \Sigma_Z \end{pmatrix} = \left[\begin{pmatrix} L' & 0 \\ I & C' \end{pmatrix} \begin{pmatrix} \Sigma_e & \Sigma_{eu} \\ \Sigma'_{eu} & \Sigma_u \end{pmatrix}^{-1} \begin{pmatrix} L & I \\ 0 & C \end{pmatrix} \right]^{-1}$$

$$= \begin{pmatrix} A & B \\ B' & D \end{pmatrix}^{-1}, \quad (A.4)$$

where, assuming that all the inverse matrices involved exist

$$A = L'(\Sigma_e^{-1} + \Sigma_e^{-1}\Sigma_{eu}\Sigma_{u|e}^{-1}\Sigma'_{eu}\Sigma_e^{-1})L, \quad B = L'(\Sigma_e^{-1} + \Sigma_e^{-1}\Sigma_{eu}\Sigma_{u|e}^{-1}\Sigma'_f)$$

and

$$D = \Sigma_e^{-1} + \Sigma_f\Sigma_{u|e}^{-1}\Sigma'_f \quad (A.5)$$

by calling $\Sigma_f = \Sigma_e^{-1}\Sigma_{eu} - C'$ and $\Sigma_{u|e} = \Sigma_u - \Sigma'_{eu}\Sigma_e^{-1}\Sigma_{eu}$. Then, inversion by blocks yields

$$\Sigma_b = (A - BD^{-1}B')^{-1}, \quad \Sigma_{bZ} = -A^{-1}B(D - B'A^{-1}B)^{-1}$$

and

$$\Sigma_Z = (D - B'A^{-1}B)^{-1}. \quad (A.6)$$

Next, the following matrix inversion lemma:

$$(E + FGH)^{-1} = E^{-1} - E^{-1}F(G^{-1} + HE^{-1}F)^{-1}HE^{-1}, \quad (A.7)$$

which holds for E and G nonsingular matrices, allows us to obtain

$$\Sigma_b = [L'C'(\Sigma_{u|e} + \Sigma'_f\Sigma_e\Sigma_f)^{-1}CL]^{-1} = (L'C'\Sigma_d^{-1}CL)^{-1}, \quad (A.8)$$

where $\Sigma_d = \mathrm{Var}(\mathbf{u} - C\mathbf{e}|X) = \Sigma_u - C\Sigma_{eu} - \Sigma'_{eu}C' + C\Sigma_e C'$. Similarly we get

$$\Sigma_Z = D^{-1} + D^{-1}B'\Sigma_b BD^{-1}$$

$$= (\Sigma_e - \Sigma_e\Sigma_f\Sigma_d^{-1}\Sigma'_f\Sigma_e) + (I - \Sigma_e\Sigma_f\Sigma_d^{-1}C)L\Sigma_b L'(I - C'\Sigma_d^{-1}\Sigma'_f\Sigma_e) \quad (A.9)$$

and

$$\Sigma_{bZ} = -\Sigma_b BD^{-1} = -\Sigma_b L'(I - C'\Sigma_d^{-1}\Sigma'_f\Sigma_e). \quad (A.10)$$

Now, from (A.3) it follows that:

$$\begin{pmatrix} \hat{\mathbf{b}} \\ E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y}) \end{pmatrix} = \begin{pmatrix} \Sigma_b & \Sigma_{bZ} \\ \Sigma'_{bZ} & \Sigma_Z \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \hat{\mathbf{Z}}_X \\ \mathbf{Y} \end{pmatrix}$$

$$= \begin{pmatrix} (\Sigma_b\alpha + \Sigma_{bZ}\gamma)\hat{\mathbf{Z}}_X + (\Sigma_b\beta + \Sigma_{bZ}\delta)\mathbf{Y} \\ (\Sigma'_{bZ}\alpha + \Sigma_Z\gamma)\hat{\mathbf{Z}}_X + (\Sigma'_{bZ}\beta + \Sigma_Z\delta)\mathbf{Y} \end{pmatrix} \quad (A.11)$$

with $\alpha = L'(\Sigma_e^{-1} + \Sigma_e^{-1}\Sigma_{eu}\Sigma_{u|e}^{-1}\Sigma_{eu}'\Sigma_e^{-1})$, $\beta = -L'\Sigma_e^{-1}\Sigma_{eu}\Sigma_{u|e}^{-1}$, $\gamma = \Sigma_e^{-1} + \Sigma_f\Sigma_{u|e}^{-1}\Sigma_{eu}'\Sigma_e^{-1}$ and $\delta = -\Sigma_f\Sigma_{u|e}^{-1}$. Then, after some tedious algebraic manipulations we get

$$\Sigma_b\alpha + \Sigma_{bZ}\gamma = \Sigma_b L'C'\Sigma_d^{-1} \quad \text{and} \quad \Sigma_b\beta + \Sigma_{bZ}\delta = -\Sigma_b L'C'\Sigma_d^{-1} \tag{A.12}$$

so that

$$\hat{\mathbf{b}} = -\Sigma_b L'C'\Sigma_d^{-1}(\mathbf{Y} - C\hat{\mathbf{Z}}_X). \tag{A.13}$$

Furthermore

$$\Sigma_{bZ}'\beta + \Sigma_Z\delta = (I + \Sigma_e\Sigma_f\Sigma_d^{-1}C)L\Sigma_b L'C'\Sigma_d^{-1} - \Sigma_e\Sigma_f\Sigma_d^{-1} \tag{A.14}$$

and, since $\alpha = L'\Sigma_e^{-1} - \beta\Sigma_{eu}'\Sigma_e^{-1}$ and $\gamma = \Sigma_e^{-1} - \delta\Sigma_{eu}'\Sigma_e^{-1}$, we also have

$$\begin{aligned}\Sigma_{bZ}'\alpha + \Sigma_Z\gamma &= (\Sigma_{bZ}'L' + \Sigma_Z)\Sigma_e^{-1} - (\Sigma_{bZ}'\beta + \Sigma_Z\delta)\Sigma_{eu}'\Sigma_e^{-1}\\ &= I + \Sigma_e\Sigma_f\Sigma_d^{-1}C - (I + \Sigma_e\Sigma_f\Sigma_d^{-1}C)L\Sigma_b L'C'\Sigma_d^{-1}C \end{aligned} \tag{A.15}$$

hence

$$\begin{aligned} E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y}) &= \hat{\mathbf{Z}}_X + [(I + \Sigma_e\Sigma_f\Sigma_d^{-1}C)L\Sigma_b L'C'\Sigma_d^{-1} - \Sigma_e\Sigma_f\Sigma_d^{-1}](\mathbf{Y} - C\hat{\mathbf{Z}}_X)\\ &= \hat{\mathbf{Z}}_X - \Sigma_e\Sigma_f\Sigma_d^{-1}(\mathbf{Y} - C\hat{\mathbf{Z}}_X) - (I + \Sigma_e\Sigma_f\Sigma_d^{-1}C)L\hat{\mathbf{b}}\\ &= (\hat{\mathbf{Z}}_X - L\hat{\mathbf{b}}) - \Sigma_e\Sigma_f\Sigma_d^{-1}[\mathbf{Y} - C(\hat{\mathbf{Z}}_X - L\hat{\mathbf{b}})] \end{aligned} \tag{A.16}$$

Finally, it is clear that

$$\begin{aligned} E(\hat{\mathbf{b}}|X) &= -\Sigma_b L'C'\Sigma_d^{-1}E(\mathbf{u} - C\mathbf{e} - CL\mathbf{b}|X)\\ &= \Sigma_b L'C'\Sigma_d^{-1}CL\mathbf{b} = \mathbf{b} \end{aligned} \tag{A.17}$$

and

$$\begin{aligned} E[E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y}) - \hat{\mathbf{Z}}_X|X] &= E(\hat{\mathbf{Z}}_X - \mathbf{Z} - L\hat{\mathbf{b}}|X) - \Sigma_e\Sigma_f\Sigma_d^{-1}E[\mathbf{Y} - C(\hat{\mathbf{Z}}_X - L\hat{\mathbf{b}})]\\ &= E[L(\mathbf{b} - \hat{\mathbf{b}}) + \mathbf{e}|X)]\\ &\quad - \Sigma_e\Sigma_f\Sigma_d^{-1}E[\mathbf{u} - C\mathbf{e} - CL(\mathbf{b} - \hat{\mathbf{b}})|X]\\ &= \mathbf{0}. \end{aligned} \tag{A.18}$$

Appendix B. Proof of the GCR when no bias is present

Let $L = 0$, then $\alpha = 0$ and $\beta = 0$ in (A.11), so that

$$\begin{aligned} E(\mathbf{Z}|\hat{\mathbf{Z}}_X, \mathbf{Y}) &= \Sigma_Z\gamma\hat{\mathbf{Z}}_X + \Sigma_Z\delta\mathbf{Y}\\ &= \Sigma_Z(\Sigma_e^{-1} + \Sigma_f\Sigma_{u|e}^{-1}\Sigma_{eu}'\Sigma_e^{-1})\hat{\mathbf{Z}}_X - \Sigma_Z\Sigma_f\Sigma_{u|e}^{-1}\mathbf{Y} \end{aligned} \tag{B.1}$$

and

$$\Sigma_Z = \Sigma_e - \Sigma_e\Sigma_f\Sigma_d^{-1}\Sigma_f'\Sigma_e = (\Sigma_e^{-1} + \Sigma_f\Sigma_{u|e}^{-1}\Sigma_f')^{-1}. \tag{B.2}$$

Appendix C. Proof of expressions (2.12) and (2.13)

We now consider the equation

$$\mathbf{Y} - \Sigma'_{eu}\Sigma_e^{-1}\hat{\mathbf{Z}}_X = (C\mathbf{Z} + \mathbf{u}) - \Sigma'_{eu}\Sigma_e^{-1}(\mathbf{Z} + \mathbf{e}) \tag{C.1}$$

that is

$$\hat{\mathbf{Y}} = (C - \Sigma'_{eu}\Sigma_e^{-1})\mathbf{Z} + \boldsymbol{\varepsilon} \quad \text{with } \boldsymbol{\varepsilon} = \mathbf{u} - \Sigma'_{eu}\Sigma_e^{-1}\mathbf{e} \tag{C.2}$$

so that $E(\boldsymbol{\varepsilon}|X) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|X) = \Sigma_{u|e}$. Then, by generalized least squares we obtain

$$E(\mathbf{Z}|X, \hat{\mathbf{Y}}) = -(\Sigma_f\Sigma_{u|e}^{-1}\Sigma'_f)^{-1}\Sigma_f\Sigma_{u|e}^{-1}\hat{\mathbf{Y}}. \tag{C.3}$$

Now (C.2) implies that

$$E[E(\mathbf{Z}|X, \hat{\mathbf{Y}}) - \mathbf{Z}|X] = E[-(\Sigma_f\Sigma_{u|e}^{-1}\Sigma'_f)^{-1}\Sigma_f\Sigma_{u|e}^{-1}\boldsymbol{\varepsilon}|X] = \mathbf{0} \tag{C.4}$$

so that

$$\begin{aligned}\mathrm{Var}[E(\mathbf{Z}|X, \hat{\mathbf{Y}}) - \mathbf{Z}] &= E[(\Sigma_f\Sigma_{u|e}^{-1}\Sigma'_f)^{-1}\Sigma_f\Sigma_{u|e}^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\Sigma_{u|e}^{-1}\Sigma'_f(\Sigma_f\Sigma_{u|e}^{-1}\Sigma'_f)^{-1}] \\ &= (\Sigma_f\Sigma_{u|e}^{-1}\Sigma'_f)^{-1}. \end{aligned} \tag{C.5}$$

References

Aoki, M., 1990. State Space Modeling of Time Series. Springer, Berlin.

Clements, M.P., Hendry, D.F., 1993. On the limitations of comparing mean square forecast errors. J. Forecasting 12, 617–637.

Clements, M.P., Hendry, D.F., 1999. Forecasting Non–Stationary Economic Time Series. MIT Press, Cambridge, MA.

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

Dagum, E.B., Cholette, P.A., Chen, Z.G., 1998. A unified view of signal extraction, benchmarking, interpolation and extrapolation of time series. Int. Statist. Rev. 66, 245–269.

Denton, F.T., 1971. Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization J. Amer. Statist. Assoc. 66, 99–102.

Doan, T., Litterman, R., Sims, C.A., 1984. Forecasting and conditional projection using realistic prior distributions. Econom. Rev. 3, 1–100.

Greene, M.N., Howrey, P., Hymans, S.H., 1986. The use of outside information in econometric forecasting. In: Kuh, E., Belsley, D.A. (Eds.), Model Reliability. MIT Press, Cambridge, MA.

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

Guerrero, V.M., 1990. Restricted ARIMA forecasts which account for parameter changes. ESTADISTICA 42, 17–31.

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

Guerrero, V.M., Berumen, E., 1998. Forecasting electricity consumption with extra-model information provided by consumers. J. Appl. Statist. 25, 283–299.

Guerrero, V.M., Peña, D., 2000. Linear combination of restrictions and forecasts in time series analysis. J. Forecasting 19, 103–122.

Harvey, A.C., Chung, C.-H., 2000. Estimating the underlying change in UK unemployment. J. Roy. Statist. Soc. A-163, 303–339.

Harvey, A.C., Stock, J.H., 1993. Estimation, smoothing, interpolation and distribution in structural time series models in continuous time. In: Phillips, P.C.B. (Ed.), Models, Methods, and Applications of Econometrics, pp. 55–70.

Judge, G.C., Griffiths, W.E., Hill, R.C., Lee, T. Ch., 1980. The Theory and Practice of Econometrics. Wiley, New York.

Lütkepohl, H., 1991. Introduction to Multiple Time Series Analysis. Springer, Berlin.

Matthews, K.G.P., Minford, A.P.L., Blackman, S.C., 1994. The quarterly liverpool model with current partial information: an exercise in forecasting J. Forecasting 13, 507–518.

Newbold, P., Granger, C.W.J., 1974. Experience with forecasting univariate time series and the combination of forecasts. J. Roy. Statist. Soc. A-137, 131–164.

Pankratz, A., 1989. Time series forecasts and extra-model information. J. Forecasting 8, 75–83.

Peña, D., 1997. Combining information in statistical modeling. Amer. Statist. 51, 326–332.

Pino, F.A., Morettin, P.A., Mentz, R.P., 1987. Modelling and forecasting linear combination of time series. Int. Statist. Rev. 3, 295–313.

Reinsel, G.C., 1993. Elements of Multivariate Time Series Analysis. Springer, Berlin.

Sandee, J., Don, F.J.H., Van der Berg, P.J.C.M., 1984. Adjustment of projections to recent observations. European Econom. Rev. 26, 153–166.

Shumway, R.H., Stoffer, D., 2000. Time Series Analysis and its Applications. Springer, Berlin.

Theil, H., 1963. On the use of incomplete prior information in regression analysis. J. Amer. Statist. Assoc. 58, 401–414.

Theil, H., 1974. Mixed estimation based on quasi-prior judgments. European Econom. Rev. 5, 33–40.

Theil, H., Goldberger, A.S., 1961. On pure and mixed statistical estimation in economics. Int. Econom. Rev. 2, 65–78.

Tiao, G.C., Box, G.E.P., 1981. Modeling multiple time series with applications. J. Amer. Statist. Assoc. 76, 802–816.

Trabelsi, A., Hillmer, S.C., 1989. A benchmarking approach to forecast combination. J. Business Econom. Statist. 7, 353–362.

Tsay, R.S., 1988. Outliers, level shifts and variance changes in time series. J. Forecasting 7, 1–20.

Wallis, K.F., 1986. Forecasting with an econometric model: the ‘ragged edge’ problem J. Forecasting 5, 1–13.

Wei, W.W.S., Abraham, B., 1981. Forecasting contemporal time series aggregates. Comm. Statist. A-10, 1335–1344.