Journal of Forecasting, Vol. 8, 215–229 (1989)

Optimal Conditional ARIMA Forecasts

VÍCTOR M. GUERRERO
Banco de México, México

ABSTRACT

An optimal univariate forecast, based on historical and additional information about the future, is obtained in this paper. Its statistical properties, as well as some inferential procedures derived from it, are indicated. Two main situations are considered explicitly: (1) when the additional information imposes a constraint to be fulfilled exactly by the forecasts and (2) when the information is only a conjecture about the future values of the series or a forecast from an alternative model. Theoretical and empirical illustrations are provided, and a unification of the existing methods is also attempted.

KEY WORDS ARIMA models Combination of forecasts Minimum mean-square error Prior information Quadratic minimization Time series

INTRODUCTION

There are many situations in which a time series analyst is faced with the problem of constraining the ARIMA forecasts of a given variable so that a condition imposed on the future values of the series is fulfilled. One such situation was considered by Vera and Guerrero (1981), where it was known that the financing granted by the Mexican bank system had to reach a particular value at the end of a future year, but the monthly stock values for that year were not restricted except by the historical behavior of the series.

Another situation put forward by Cholette (1982) indicated that a restriction on the annual level of a series may be contemplated as a scenario or as a prediction from an econometric model, but the details about the sub-annual path which would be followed in realizing that prediction are not provided. Such restrictions on the future values of the time series under study are usually given in the form of linear combinations (e.g. averages, differences between two particular values or isolated values) to be fulfilled by the forecasts.

In this paper an approach is adopted which combines such linear restrictions with the forecasts obtained from a univariate ARIMA model. The optimal conditional forecast here derived has appeared in an implicit form in the work of Doan *et al.* (1983) and Cholette (1982). It appeared more explicitly in an unpublished manuscript by Pérez-Porrúa (1984), but in none of these papers are the statistical properties of the optimal forecast emphasized nor is the appropriateness of the procedure indicated. Here we derived such properties and procedures, which serve in particular as a guide for deciding when the formulas can be appropriately used.

0277–6693/89/030215–15$07.50 *Received May 1987*

Revised February 1988

We have also attempted to provide a unified framework by using the Box–Jenkins notation, which allowed us to reconcile the apparently different methodologies, particularly those of Cholette and Pérez-Porrúa, which are shown to be strictly equivalent.

Doan *et al.* (1983), using a general multivariate setting, derived conditional forecasts by an application of the Law of Iterated Projections, but the statistical aspects of their procedure are not emphasized, and we can easily overlook those aspects by paying attention to the mathematical details of the optimization problem involved. Cholette (1982) makes use of the autoregressive representation of the ARIMA model and solves a quadratic minimization problem, which is a generalization of the case considered by Doan *et al.* (1983) in the sense that the restriction need not be taken as a given fact but can be considered as conjecture. The solution to this problem yields a conditional forecast. However, this solution is difficult to implement in practice for long series, due to the computational burden. Again, no statistical properties or inferential procedures are provided by the author. Finally, Pérez-Porrúa (1984) solved another general problem, which will be shown to be identical to that posed by Cholette, but now using the moving-average form of the ARIMA model and with essentially the same tools employed by Doan *et al.* (1983).

The formulation presented here is more statistically oriented, since the properties of the forecasts, as well as some statistics, are derived in order to be able to judge the appropriateness of the procedures suggested. In the following section a framework is provided for obtaining optimal (minimum mean-square error, linear) conditional forecasts. The third section considers the problem of testing whether the restriction is compatible with the historical data and a statistic is derived. In the fourth section some theoretical illustrations are given in order to gain some insight into the calculations involved, and the results are seen to be very intuitively appealing. The generalization considered by Pérez-Porrúa (1984) is presented and justified in the fifth section, in which a test statistic is also derived as an aid for choosing the variance–covariance matrix involved in the generalization. This section also shows the formal equivalence between the results obtained by Pérez-Porrúa (1984) and those of Cholette (1982), and a theoretical illustration of the generalized case is also given here. The sixth section is devoted to applications in which empirical data are used only for illustrative purposes and the final section presents some conclusions.

DERIVATION OF THE OPTIMAL CONDITIONAL FORECASTS

Let $\{Z_t\}$ be a time series which is observed during $t = 1, \ldots, N$ and for which an ARIMA model can be constructed. The moving-average representation of the series is

$$Z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \tag{1}$$

where the variables a_t form a sequence of random shocks with mean zero and variance σ_a^2 and the weights ψ_j are known constants.

Box and Jenkins (1970) show that the optimal (linear with minimum mean-square error) forecast of Z_{N+h}, given the vector of observations $\mathbf{Z}_0 = (Z_1, \ldots, Z_N)'$ is

$$E(Z_{N+h} \mid \mathbf{Z}_0) = \sum_{j=h}^{\infty} \psi_j a_{N+h-j}, h = 1, \ldots, H \tag{2}$$

with its corresponding forecast error given by

$$Z_{N+h} - E(Z_{N+h} \mid \mathbf{Z}_0) = \sum_{j=0}^{h-1} \psi_j a_{N+h-j} \tag{3}$$

In matrix notation we have

$$\mathbf{Z}_F - E(\mathbf{Z}_F \mid \mathbf{Z}_0) = \boldsymbol{\psi} \mathbf{a}_F \tag{4}$$

with

$$\mathbf{Z}_F = (Z_{N+1}, \ldots, Z_{N+H})', \mathbf{a}_F = (a_{N+1}, \ldots, a_{N+H})'$$

and $\boldsymbol{\psi}$ is the $H \times H$ dimensional matrix

$$\boldsymbol{\psi} = \begin{bmatrix} \psi_0 & 0 & \ldots & 0 \\ \psi_1 & \psi_0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ \psi_{H-1} & \psi_{H-2} & \ldots & \psi_0 \end{bmatrix} \text{with } \psi_0 = 1$$

where

$$E(\mathbf{a}_F \mid \mathbf{Z}_0) = \mathbf{0} \quad \text{and} \quad E(\mathbf{a}_F \mathbf{a}_F' \mid \mathbf{Z}_0) = \sigma_a^2 I.$$

Let us assume that the additional information is given as the linear combination

$$\mathbf{Y} = C\mathbf{Z}_F \tag{5}$$

with C an $m \times H$ matrix ($m \leqslant H$) of constants. Then by equation (4), we have

$$\mathbf{Y} = CE(\mathbf{Z}_F \mid \mathbf{Z}_0) + C\boldsymbol{\psi}\mathbf{a}_F \tag{6}$$

The optimal conditional forecast

In order to obtain the optimal forecast which incorporates condition (6) there must exist an $H \times m$ constant matrix A such that

$$\begin{aligned} \tilde{\mathbf{Z}}_F &= A\mathbf{Y} \\ &= AC[E(\mathbf{Z}_F \mid \mathbf{Z}_0) + \boldsymbol{\psi}\mathbf{a}_F] \end{aligned} \tag{7}$$

which, by unbiasedness of the forecast error given $\mathbf{Z}_0$, must satisfy

$$\begin{aligned} \mathbf{0} &= E[(\tilde{\mathbf{Z}}_F - \mathbf{Z}_F) \mid \mathbf{Z}_0] \\ &= (AC - I)E(\mathbf{Z}_F \mid \mathbf{Z}_0) \end{aligned} \tag{8}$$

so that

$$\tilde{\mathbf{Z}}_F - \mathbf{Z}_F = (AC - I)\boldsymbol{\psi}\mathbf{a}_F \tag{9}$$

and

$$\operatorname{cov}[(\tilde{\mathbf{Z}}_F - \mathbf{Z}_F) \mid \mathbf{Z}_0] = \sigma_a^2(AC\boldsymbol{\psi}\boldsymbol{\psi}'C'A' - AC\boldsymbol{\psi}\boldsymbol{\psi}' - \boldsymbol{\psi}\boldsymbol{\psi}'C'A' + \boldsymbol{\psi}\boldsymbol{\psi}') \tag{10}$$

Therefore for $\tilde{\mathbf{Z}}_F$ to have minimum mean-square error, A must be chosen as the minimizer of the generalized variance, defined as

$$\operatorname{var}_F = \operatorname{tr}\{\operatorname{cov}[(\tilde{\mathbf{Z}}_F - \mathbf{Z}_F) \mid \mathbf{Z}_0]\} \tag{11}$$

Then A is found by solving $0 = (\partial \operatorname{var}_F / \partial A) \mid_{\hat{A}}$; that is,

$$\hat{A} = \boldsymbol{\psi}\boldsymbol{\psi}'C'(C\boldsymbol{\psi}\boldsymbol{\psi}'C')^{-1} \tag{12}$$

From equations (7) and (8) the optimal conditional forecast becomes

$$\hat{\mathbf{Z}}_F = E(\mathbf{Z}_F \mid \mathbf{Z}_0) + \hat{A}C\psi\mathbf{a}_F \tag{13}$$

Thus, using equation (6), it follows that

$$\begin{aligned}\hat{\mathbf{Z}}_F &= E(\mathbf{Z}_F \mid \mathbf{Z}_0) + \hat{A}[\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0] \\ &= E(\mathbf{Z}_F \mid \mathbf{Z}_0) + \psi\psi' C'(C\psi\psi' C')^{-1}[\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0]\end{aligned} \tag{14}$$

Alternative approaches

An alternative approach for obtaining the optimal conditional forecast is based on quadratic minimization. That is, let us consider the problem

$$\min_{\mathbf{Z}_F}\{\mathbf{a}_F'\mathbf{a}_F\} \text{ subject to } \mathbf{Y} = C\mathbf{Z}_F \tag{15}$$

Now, by equation (4) we know

$$\mathbf{a}_F = \psi^{-1}[\mathbf{Z}_F - E(\mathbf{Z}_F \mid \mathbf{Z}_0)] \tag{16}$$

Then we construct the function

$$Q = [\mathbf{Z}_F - E(\mathbf{Z}_F \mid \mathbf{Z}_0)]'(\psi\psi')^{-1}[\mathbf{Z}_F - E(\mathbf{Z}_F \mid \mathbf{Z}_0)] + 2\boldsymbol{\lambda}'(\mathbf{Y} - C\mathbf{Z}_F) \tag{17}$$

with $\boldsymbol{\lambda}$ a vector of Lagrange multipliers. Thus from $0 = (\partial Q/\partial \mathbf{Z}_F)|_{\hat{\mathbf{Z}}_f,\hat{\lambda}}$ we obtain

$$\begin{aligned}\hat{\mathbf{Z}}_F &= \mathbf{E}(\mathbf{Z}_F \mid \mathbf{Z}_0) + \psi\psi' C' \ \hat{\boldsymbol{\lambda}} \\ &= E(\mathbf{Z}_F \mid \mathbf{Z}_0) + \psi\psi' C'(C\psi\psi' C')^{-1}[\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0)]\end{aligned} \tag{18}$$

which is identical to equation (14).

Another way of obtaining $\hat{\mathbf{Z}}_F$ is the following. Let us consider the problem

$$\min_{E(\mathbf{a}_F \mid \mathbf{Z}_0)} \{E(\mathbf{a}_F \mid \mathbf{Z}_0)'E(\mathbf{a}_F \mid \mathbf{Z}_0\} \text{ subject to } C\psi E(\mathbf{a}_F \mid \mathbf{Z}_0) = \mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0) \tag{19}$$

which is solved by Lagrangian minimization and yields the solution

$$E(\mathbf{a}_F \mid \mathbf{Z}_0, \mathbf{Y}) = \psi' C'(C\psi\psi' C')^{-1}[\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0)] \tag{20}$$

Therefore, substituting this for $\mathbf{a}_F$ in equation (4) we obtain equation (14) again. This alternative method of solving the problem was cited by Doan *et al.* (1983).

A STATISTICAL TEST OF THE HYPOTHESIS $\mathbf{Y} = C\mathbf{Z}_F$

Once an ARIMA model has been constructed for $\{Z_t\}$ we can substitute the ψ matrix by its estimate and, given the additional information $\mathbf{Y}$, we can obtain the optimal conditional forecast. This, however, relies on the assumption that $\mathbf{Y}$ is of the form of equation (5), where each element of $\mathbf{Z}_F$ is generated by process (1). To decide whether $\mathbf{Y}$ can be considered consistent with the historical data, we should use a statistical test such as that suggested in this section.

Let us assume that representation (1) holds valid for $t = 1, \ldots, N + H$ and that each a_t is normal. Then we have

$$E(\mathbf{a}_F \mid \mathbf{Z}_0) = \mathbf{0} \quad \text{and} \quad \operatorname{cov}(\mathbf{a}_F \mid \mathbf{Z}_0) = \sigma_a^2 I \tag{21}$$

Thus, given $\mathbf{Z}_0$,

$$\begin{aligned} \mathbf{e}_Y &= \mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0) \\ &= C\psi\mathbf{a}_F \sim N_m(\mathbf{0}, \sigma_a^2 C\psi\psi' C') \end{aligned} \tag{22}$$

Hence, if σ_a^2 were known,

$$K_1 = \mathbf{e}_Y'(C\psi\psi' C')^{-1}\mathbf{e}_Y/\sigma_a^2 \sim \chi_m^2 \tag{23}$$

However, as it always happens in practice that

$$\hat{\sigma}_a^2 = \mathbf{a}_0'\mathbf{a}_0/(N-k) \tag{24}$$

with $\mathbf{a}_0 = (a_1, \ldots, a_N)'$ and k being the number of parameters in the ARIMA model, then

$$K_2 = \mathbf{e}_Y'(C\psi\psi' C')^{-1}\mathbf{e}_Y/(m\hat{\sigma}_a^2) \sim F_{m,N-k} \tag{25}$$

It follows that a test of the hypothesis that $\mathbf{Y} = C\mathbf{Z}_F$ can be performed by calculating the statistic

$$\hat{K}_2 = [\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0)]'(C\psi\psi' C')^{-1}[\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0)]/(m\hat{\sigma}_a^2) \tag{26}$$

and comparing its value with tabled values of an F-distribution with m and $N-k$ degrees of freedom, for a desired α significance level.

We should note however, that equation (25) holds only as an approximation when (as is the usual case) ψ is estimated from the historical data. Therefore instead of $\hat{K}_2$ we would rather use

$$\hat{K}_1 = [\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0)]'(C\psi\psi' C')^{-1}[\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0)]/\hat{\sigma}_a^2 \tag{27}$$

as an approximate chi-square statistic with m degrees of freedom.

If the calculated $\hat{K}_1$(or $\hat{K}_2$) statistic shows no evidence against the hypothesis that $\mathbf{Y}$ was generated by the original process we can be confident in using the optimal conditional forecasts given by equation (14). Furthermore, confidence bands for $\mathbf{Z}_F$, given $\mathbf{Z}_0$ and $\mathbf{Y}$, can be derived from the distribution of the forecast error

$$\mathbf{Z}_F - \hat{\mathbf{Z}}_F = (I - \hat{A}C)\psi\mathbf{a}_F = \mathbf{e}_F \tag{28}$$

which has a normal distribution with

$$E(\mathbf{e}_F \mid \mathbf{Z}_0, \mathbf{Y}) = \mathbf{0} \quad \text{and} \quad \operatorname{cov}(\mathbf{e}_F \mid \mathbf{Z}_0, \mathbf{Y}) = \sigma_a^2\psi\psi'(I - \hat{A}C)' \tag{29}$$

An important fact related to $\hat{\mathbf{Z}}_F$ is the following. From equation (4) we know that the forecast error, given only the historical data, has

$$\operatorname{cov}(\psi\mathbf{a}_F \mid \mathbf{Z}_0) = \sigma_a^2\psi\psi' \tag{30}$$

Therefore the covariance matrix for the forecasts $E(\mathbf{Z}_F \mid \mathbf{Z}_0)$ exceeds the covariance matrix of $\hat{\mathbf{Z}}_F$ by

$$\sigma_a^2\hat{A}C\psi\psi' = \sigma_a^2\psi\psi' C'(C\psi\psi' C')^{-1}C\psi\psi' \tag{31}$$

which is positive semi-definite. Hence the optimal forecasts conditional on the information $\mathbf{Z}_0$ and $\mathbf{Y}$ will be more precise than those obtained using only $\mathbf{Z}_0$.

If, on the other hand, the hypothesis $\mathbf{Y} = C\mathbf{Z}_F$ is rejected by the empirical evidence we may consider the alternative suggested later in this paper.

SOME ILLUSTRATIVE THEORETICAL EXAMPLES

In order to see how equation (14) works let us consider the following case:

$$Z_t \sim \text{ARMA}(1,1) \text{ with } E(Z_t)=0 \quad \text{and} \quad C=(0,0,0,1)$$

Here we have

$$\begin{aligned} Z_t &= \phi Z_{t-1} + a_t - \theta a_{t-1} \\ &= \sum_{j=1}^{\infty} \phi^{j-1}(\phi-\theta)a_{t-j} \end{aligned} \tag{32}$$

Therefore

$$E(Z_{N+h} \mid \mathbf{Z}_0) = \phi^{h-1}\hat{Z}_N(1), h = 1, \ldots, 4 \text{ with } \hat{Z}_N(1) = \phi Z_N - \theta a_N \tag{33}$$

so that

$$E(\mathbf{Z}_F \mid \mathbf{Z}_0) = \hat{Z}_N(1)(1, \phi, \phi^2, \phi^3)' \tag{34}$$

Now, let $Y = Z_{N+4}$ be given as additional information. Thus

$$Y - CE(\mathbf{Z}_F \mid \mathbf{Z}_0) = Y - \phi^3 \hat{Z}_N(1) \tag{35}$$

and

$$\hat{A} = l^{-1}(\phi^2(\phi-\theta), \phi(\phi-\theta)[1+\phi(\phi-\theta)], (\phi-\theta)[1+\phi(\phi-\theta)+\phi^3(\phi-\theta)], l)' \tag{36}$$

with

$$l = 1 + (\phi-\theta)^2 + \phi^2(\phi-\theta)^2 + \phi^4(\phi-\theta)^2 \tag{37}$$

Then the optimal conditional forecast, given $\mathbf{Z}_0$ and $\mathbf{Y}$, becomes

$$\hat{\mathbf{Z}}_F = l^{-1} \begin{bmatrix} (l - \phi^5(\phi-\theta))\hat{Z}_N(1) + \phi^2(\phi-\theta)Y \\ (l - \phi^3(\phi-\theta)[1+\phi(\phi-\theta)])\phi\hat{Z}_N(1) + \phi(\phi-\theta)[1+\phi(\phi-\theta)]Y \\ (l - \phi(\phi-\theta)[1+\phi(\phi-\theta)+\phi^3(\phi-\theta)])\phi^2\hat{Z}_N(1) + (\phi-\theta) \\ \qquad [1+\phi(\phi-\theta)+\phi^3(\phi-\theta)]Y \\ lY \end{bmatrix} \tag{38}$$

and the calculated statistic

$$\hat{K}_1 = \hat{K}_2 = [Y - \phi^3\hat{Z}_N(1)]^2/(l\hat{\sigma}_a^2) \tag{39}$$

should be compared with tables of χ_1^2 (or with tables of an F-distribution with 1 and $N-2$ degrees of freedom).

Some specific cases

By assigning specific values to the parameters ϕ and θ the following cases are obtained

(1) $Z_t \sim$ white noise, $\hat{Z}_N(1) = 0$
$\hat{\mathbf{Z}}_F = (0, 0, 0, Y)'$
$\hat{K}_1 = (Y/\hat{\sigma}_a)^2$

(2) $Z_t \sim \text{MA}(1), \hat{Z}_N(1) = -\theta a_N$
$\hat{\mathbf{Z}}_F = (-\theta a_N, 0, -\theta Y/(1+\theta^2), Y)'$
$\hat{K}_1 = Y^2/[(1+\theta^2)\hat{\sigma}_a^2]$

(3) $Z_t \sim \mathrm{AR}(1), \hat{Z}_N(1) = \phi Z_N$

$$\hat{\mathbf{Z}}_F = l^{-1}\begin{bmatrix}(1+\phi^2+\phi^4)\phi Z_N + \phi^3 Y \\ (1+\phi^2)\phi^2 Z_N + (1+\phi^2)\phi^2 Y \\ \phi^3 Z_N + (1+\phi^2+\phi^4)\phi Y \\ lY\end{bmatrix}, \quad l = 1+\phi^2+\phi^4+\phi^6$$

$$\hat{K}_1 = (1-\phi^2)(Y-\phi^4 Z_N)^2/[(1-\phi^8)\hat{\sigma}_a^2]$$

(4) $Z_t \sim \mathrm{IMA}(1,1), \hat{Z}_N(1) = Z_N - \theta a_N$

$$\hat{\mathbf{Z}}_F = l^{-1}\begin{bmatrix}[l-(1-\theta)]\hat{Z}_N(1) + (1-\theta)Y \\ [l-(1-\theta)(2-\theta)]\hat{Z}_N(1) + (1-\theta)(2-\theta)Y \\ [l-(1-\theta)(3-2\theta)]\hat{Z}_N(1) + (1-\theta)(3-2\theta)Y \\ lY\end{bmatrix}, \quad l = 1+3(1-\theta)^2$$

$$\hat{K}_1 = (Y - Z_N - \theta a_N)^2/(\sigma_a^2[1+3(1-\theta)^2])$$

(5) $Z_t \sim \mathrm{ARIMA}(0,1,0), \hat{Z}_N(1) = Z_N$

$$\hat{\mathbf{Z}}_F = \tfrac{1}{4}(3Z_N + Y, 2Z_N + 2Y, Z_N + 3Y, 4Y)'$$

$$\hat{K}_1 = (Y - Z_N)^2/(2\hat{\sigma}_a)^2$$

(6) $Z_t = \log(X_t) \sim \mathrm{ARIMA}(0,1,0)$

$$\begin{aligned}\hat{X}_N(1) &\doteq \exp\{E[\log(X_{N+1}) \mid \mathbf{X}_0]\} \\ &= \exp[\hat{Z}_N(1)] \\ &= X_N\end{aligned}$$

$$\hat{\mathbf{X}}_F = \left(X_N\left(\frac{e^Y}{X_N}\right)^{1/4}, X_N\left(\frac{e^Y}{X_N}\right)^{1/2}, X_N\left(\frac{e^Y}{X_N}\right)^{3/4}, e^Y\right)'$$

$$\hat{K}_1 = [Y - \log(X_N)]^2/(2\hat{\sigma}_a)^2$$

Some comments about these specific situations follow. In (1) the lack of an autocorrelation structure is seen in that the process adjusts itself to the additional information with a jump, as if it were only another random shock with a known value. (2) allows us to see how the memory of the process enters in the forecast one period ahead of the last observation and one period behind the given value. (3) shows that the optimal conditional forecasts are weighted averages of the unconditional forecast and the value of Y, but the weights do not add up to one. In (4) we can appreciate how the forecasts given by the exponential smoothing method should be combined with the additional information. (5) corresponds to a random walk model which is seen to adjust to the additional information by means of a straight-line interpolation. Finally, (6) is also a random walk model for the series expressed in logarithms (this particular situation was considered earlier by Vera and Guerrero, 1981, for a series with an approximately constant rate of growth).

A GENERALIZATION

The work of Pérez-Porrúa (1984) was focused to solve the problem in the second section of this paper, with the extra assumption that the additional information may be uncertain. Pérez-

Porrúa develops an optimal forecast assuming that

$$\mathbf{Y} - C\mathbf{Z}_F = \mathbf{u} \tag{40}$$

with $\mathbf{u}$ a random vector satisfying

$$\mathbf{u} \sim N_H(\mathbf{0}, U), E(\mathbf{Z}_0\mathbf{u}') = E(\mathbf{a}_F\mathbf{u}') = 0 \tag{41}$$

Then, by a method similar to that used by Doan *et al.* (1983), Pérez-Porrúa obtains the expression (see Appendix)

$$\begin{aligned}\hat{\mathbf{Z}}_{F,U} &= E(\mathbf{Z}_F \mid \mathbf{Z}_0, \mathbf{Y}) \\ &= E(\mathbf{Z}_F \mid \mathbf{Z}_0) + \hat{A}_U[\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0)]\end{aligned} \tag{42}$$

where

$$\hat{A}_U = \psi\psi' C'(C\psi\psi' C' + U/\sigma_a^2)^{-1} \tag{43}$$

which are the generalized counterparts of expressions (14) and (13), respectively.

Within the framework of the present paper, assumption (40) can now be justified by means of the statistical test suggested earlier, since a rejection of the hypothesis $\mathbf{Y} = C\mathbf{Z}_F$ might lead to the inclusion of the random term $\mathbf{u}$. However, there could be some practical situations in which including the random term could be deemed appropriate on *a priori* grounds, e.g. when the additional information $\mathbf{Y}$ was obtained from an econometric model in which the data employed were annual and the ARIMA model was constructed for a monthly series $\{Z_t\}$. This situation has been cited by Cholette (1982), and, in fact, it corresponds to combining forecasts from two different models.

SELECTING THE MATRIX U

In the present generalization we need to be able to specify the matrix U appearing in equation (41). This can be done either by using the same external source of information (e.g. an econometric model) from which $\mathbf{Y}$ was obtained or the internal evidence provided by the data. For the latter case we propose the following method. Since

$$\mathbf{e}_{Y,U} = \mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0) = C\psi\mathbf{a}_F + \mathbf{u} \tag{44}$$

then $\mathbf{e}_{Y,U}$ is normally distributed with

$$E(\mathbf{e}_{Y,U} \mid \mathbf{Z}_0) = \mathbf{0},\ \operatorname{cov}(\mathbf{e}_{Y,U} \mid \mathbf{Z}_0) = \sigma_a^2(C\psi\psi' C' + U) \tag{45}$$

and

$$K_U = \mathbf{e}_{Y,U}'(C\psi\psi' C' + U/\sigma_a^2)^{-1}\mathbf{e}_{Y,U}/\sigma_a^2 \sim \chi_m^2 \tag{46}$$

Therefore, for N large enough, we can calculate the statistic

$$\hat{K}_U = [\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0)]'(C\psi\psi' C' + U/\hat{\sigma}_a^2)^{-1}[\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0)]/\hat{\sigma}_a^2 \tag{47}$$

which can be used for specifying the matrix U. If it happens that

$$\hat{K}_U < \chi_m^2(\alpha) \tag{48}$$

with $\chi_m^2(\alpha)$ the upper α percentage point of the chi-square distribution with m degrees of freedom, then the hypothesis

$$\mathbf{Y} - C\mathbf{Z}_F = \mathbf{u} \quad \text{with} \quad \mathbf{u} \sim N(\mathbf{0}, U) \tag{49}$$

is not rejected by the data.

Once the matrix U is given, either on *a priori* grounds or by using the suggested statistic $\hat{K}_U$, we can obtain the optimal conditional forecast $\hat{\mathbf{Z}}_{F,U}$ by means of equations (42) and (43). Going one step further, we can also make an inference based on these forecasts, using the fact that, given $\mathbf{Z}_0$ and $\mathbf{Y}$,

$$\begin{aligned} \mathbf{Z}_F - \hat{\mathbf{Z}}_{F,U} &= (I - \hat{A}_U C)(\psi \mathbf{a}_F + \mathbf{u}) \\ &\sim N(\mathbf{0}, \sigma_a^2 \psi\psi'(I - \hat{A}_U C)') \end{aligned} \tag{50}$$

The quadratic minimization approach

Here we briefly interpret the steps followed by Cholette (1982) to obtain a conditional forecast by quadratic minimization and show the equivalence between this solution and the previous one.

Let us assume that Z_t admits the autoregressive representation

$$a_t = Z_t - \sum_{i=1}^{t-1} \pi_i Z_{t-i} \quad \text{for} \quad t = 1, \ldots, N \tag{51}$$

where $\pi_1, \ldots, \pi_{t-1}$ are known constants. Then

$$Z_{N+h} = \sum_{i=1}^{N+h-1} \pi_i Z_{N+h-i} + a_{N+h} \tag{52}$$

so that, for $h = 1, \ldots, H$,

$$E(Z_{N+h} \mid \mathbf{Z}_0) = \pi_1 E(Z_{N+h-1} \mid \mathbf{Z}_0) + \cdots + \pi_{h-1} E(Z_{N+1} \mid \mathbf{Z}_0) + \pi_h Z_N + \cdots + \pi_{N+h-1} Z_1 \tag{53}$$

Now, from equation (4) we know that

$$\mathbf{a}_F = \pi[\mathbf{Z}_F - E(\mathbf{Z}_F \mid \mathbf{Z}_0)] \text{ with } \pi = \psi^{-1} \tag{54}$$

where

$$\begin{aligned} \pi E(\mathbf{Z}_F \mid \mathbf{Z}_0) &= \begin{bmatrix} 1 & 0 & \ldots & 0 \\ -\pi_1 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ -\pi_{H-1} & -\pi_{H-2} & \ldots & 1 \end{bmatrix} \begin{bmatrix} E(Z_{N+1} \mid \mathbf{Z}_0) \\ E(Z_{N+2} \mid \mathbf{Z}_0) \\ \vdots \\ E(Z_{N+H} \mid \mathbf{Z}_0) \end{bmatrix} \\ &= \begin{bmatrix} \pi_N & \pi_{N-1} & \ldots & \pi_1 \\ \pi_{N+1} & \pi_N & \ldots & \pi_2 \\ \vdots & \vdots & & \vdots \\ \pi_{N+H-1} & \pi_{N+H-2} & \ldots & \pi_H \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{bmatrix} \\ &= \pi_* \mathbf{Z}_0 \end{aligned} \tag{55}$$

The problem, as posed by Cholette (1982) was

$$\min_{\mathbf{Z}_f} \{ \mathbf{a}_F' \mathbf{a}_F + (\mathbf{Y} - C\mathbf{Z}_F)' \mathbf{G} (\mathbf{Y} - C\mathbf{Z}_F) \} \tag{56}$$

where G is an $m \times m$ diagonal matrix of weights to be chosen by the analyst. This problem can be restated in terms of $\mathbf{Z}_F$ only by noting that

$$\begin{aligned} \mathbf{a}_F'\mathbf{a}_F &= [\mathbf{Z}_F - E(\mathbf{Z}_F \mid \mathbf{Z}_0)]'\pi'\pi[\mathbf{Z}_F - E(\mathbf{Z}_F \mid \mathbf{Z}_0)] \\ &= (\pi\mathbf{Z}_F - \pi_*\mathbf{Z}_0)'(\pi\mathbf{Z}_F - \pi_*\mathbf{Z}_0) \end{aligned} \tag{57}$$

and the solution becomes

$$\begin{aligned} \hat{\mathbf{Z}}_{F,G} &= (\pi'\pi + C'GC)^{-1}(\pi'\pi_*\mathbf{Z}_0 + C'G\mathbf{Y}) \\ &= [(\pi'\pi + C'GC)^{-1}\pi'\pi_*, (\pi'\pi + C'GC)^{-1}C'G]\begin{bmatrix}\mathbf{Z}_0\\ \mathbf{Y}\end{bmatrix} \end{aligned} \tag{58}$$

This equation appears as expression (9) in Cholette's paper, by recognizing $P_1 = \pi$, $B = C$, $P_2 = -\pi_*$ and $X = \mathbf{Z}_0$. A minor point is that Cholette unnecessarily restricts the sum that appears in equation (52) up to the Nth term, a fact noticeable in that the π_* matrix in equation (55) has all its elements below the diagonal term π_N equal to zero. This fact might affect the results when moving-average parameters are present, since not all the available information would then be used.

Two difficulties related to Cholette's solution (58) are first, handling the $(N+H)\times(N+H)$ matrix π_* and second, interpreting (and specifying) the matrix of weights G. In order to overcome these difficulties, let us note that equation (58) can be rewritten as

$$\begin{aligned} \hat{\mathbf{Z}}_{F,G} &= (\pi'\pi + C'GC)^{-1}[\pi'\pi E(\mathbf{Z}_F \mid \mathbf{Z}_0) + C'G\mathbf{Y}] \\ &= E(\mathbf{Z}_F \mid \mathbf{Z}_0) + \hat{A}_G[\mathbf{Y} - CE(\mathbf{Z}_F \mid \mathbf{Z}_0)] \end{aligned} \tag{59}$$

with

$$\hat{A}_G = (\pi'\pi + C'GC)^{-1}C'G \tag{60}$$

which is of the form of equation (42). Furthermore, algebraic manipulation of equation (60) yields the following:

$$\begin{aligned} \hat{A}_G &= [I + (\pi'\pi)^{-1}C'GC]^{-1}(\pi'\pi)^{-1}C'G \\ &= (I + \psi\psi'C'GC)^{-1}\psi\psi'C'G \\ &= \{I - \psi\psi'C'G[(G^{-1} + C\psi\psi'C')G]^{-1}C\}\psi\psi'C'G \\ &= \psi\psi'C'G - \psi\psi'C'(G^{-1} + C\psi\psi'C')^{-1}C\psi\psi'C'G \\ &= \psi\psi'C'G - \psi\psi'C'[I - (G^{-1} + C\psi\psi'C')^{-1}G^{-1}]G \\ &= \psi\psi'C'(C\psi\psi'C' + G^{-1})^{-1} \end{aligned} \tag{61}$$

which is of the form of equation (43) with $G^{-1} = U/\sigma_a^2$.

Therefore we do not have to handle the matrices π and π_* to obtain $\hat{\mathbf{Z}}_{F,G}$ but only the $H \times H$ matrix ψ. Second, since G equals the inverse of the variance–covariance matrix U/σ_a^2 appearing in equation (43) we have a method of interpreting and specifying G in practice.

A theoretical example

As in the preceding section, let us now present a theoretical illustration which will allow us to gain some insight into the calculations. Let us consider the case in which

$$Z_t \sim \text{ARMA}(1,1),\ E(Z_t) = 0,\ C = (0,0,0,1) \text{ and } U = \sigma_u^2$$

Thus we have $\hat{Z}_N(1)$ as in equation (33), $\hat{A}_U = \hat{A}l(l + \sigma_u^2/\sigma_a^2)^{-1}$ with $\hat{A}$ and l as in equations (36) and (37) and

$$\hat{\mathbf{Z}}_{F,U} = (l + \sigma_u^2/\sigma_a^2)^{-1} \begin{bmatrix} [l + \sigma_u^2/\sigma_a^2 - \phi^5(\phi - \theta)]\hat{Z}_N(1) + \phi^2(\phi - \theta)Y \\ (l + \sigma_u^2/\sigma_a^2 - \phi^3(\phi - \theta)[1 + \phi(\phi - \theta)])\phi\hat{Z}_N(1) \\ \qquad + \phi(\phi - \theta)[1 + \phi(\phi - \theta)]Y \\ (l + \sigma_u^2/\lambda_a^2 - \phi(\phi - \theta)[1 + \phi(\phi - \theta) + \phi^3(\phi - \theta)])\phi^2\hat{Z}_N(1) \\ \qquad + (\phi - \theta)[1 + \phi(\phi - \theta) + \phi^3(\phi - \theta)]Y \\ (\sigma_u^2/\sigma_a^2)\phi^3\hat{Z}_N(1) + (1 - \sigma_u^2/\sigma_a^2)Y \end{bmatrix} \tag{62}$$

The test statistic becomes

$$\hat{K}_U = [Y - \phi^3\hat{Z}_N(1)]^2/(l\hat{\sigma}_a^2 + \sigma_u^2) \tag{63}$$

which should be compared with tables of $\chi_1^2(\alpha)$ at the desired α significance level.

Now, by studying equation (51) we conclude that the conjectured value Y will not be reached exactly unless $\sigma_u^2 = 0$, which takes us back to the original case considered in the second section of this paper. Further, expression (48) indicates that we should choose σ_u^2 in such a way that the condition

$$\frac{[Y - \phi^3\hat{Z}_N(1)]^2}{\chi_1^2(\alpha)} - l\hat{\sigma}_a^2 < \sigma_u^2 \tag{64}$$

is satisfied for a given α value.

SOME ILLUSTRATIVE EMPIRICAL EXAMPLES

To provide a numerical illustration of the suggested procedures we shall consider the data on the annual percentage rate of growth of the Index of Manufacturing Production (IMP) for Mexico, denoted by $\{R_t^{\mathrm{IMP}}\}$. For this time series, a model of the form

$$\nabla R_t^{\mathrm{IMP}} = (1 - \theta B)(1 - \Theta B^{12})a_t$$

was constructed following the Box–Jenkins strategy. The estimated parameters obtained with data from January 1975 to September 1986 ($N = 129$) became

$$\hat{\theta} = 0.1031, \ \hat{\Theta} = 0.8111 \quad \text{and} \quad \hat{\sigma}_a^2 = 0.9313$$

so that the weights $\psi_1, \psi_2, \ldots$ are estimated by

$$\hat{\psi}_1 = \cdots = \hat{\psi}_{11} = 0.896876, \ \hat{\psi}_{12} = 0.085819, \ \hat{\psi}_{13} = \hat{\psi}_{14} = 0.169458$$

Conditional forecasts with only one restriction

We shall assume the interest lies in obtaining a scenario (conditional forecasts) for the year 1987 based on the restriction that the rate of growth of IMP reaches 7% in December 1987. Here we have $H = 15$, $m = 1$, $C = (0\ 0 \ldots 0\ 1)$ and $Y = 7$. We would like to apply the procedure presented in the second section of this paper (say, procedure A), unless there were evidence that the restriction is not consistent with the historical data. Therefore it is important to perform the test suggested in the third section. This we did, and obtained $\hat{K}_1 = \hat{K}_2 = 5.781$ in such a way that the restriction cannot be considered consistent with the ARIMA model at the 5% significant level. Therefore the procedure of the previous section (say, procedure A_U) should be employed.

Table 1. Forecasts for the rate of growth of IMP: $Y = 7$

	ARIMA		Procedure A		Procedure $A_U (U \doteq 4.5)$	
Month	Forecasts	Std. error	Forecasts	Std. error	Forecasts	Std. error
Oct. 1986	−3.72	0.931	−3.60	0.930	−3.64	0.930
Nov.	−3.66	1.251	−3.43	1.247	−3.51	1.249
Dec.	−3.55	1.504	−3.27	1.500	−3.37	1.501
Jan. 1987	−1.31	1.721	−0.40	1.679	−0.71	1.693
Feb.	−2.04	1.913	−0.56	1.811	−1.07	1.846
March	−2.94	2.087	−0.89	1.904	−1.59	1.969
April	−2.23	2.248	0.39	1.965	−0.51	2.067
May	−1.82	2.398	1.38	1.996	0.29	2.143
June	−1.07	2.539	2.70	1.998	1.40	2.199
July	−0.71	2.673	3.63	1.971	2.14	2.238
Aug.	−0.19	2.801	4.72	1.915	3.04	2.259
Sep.	−0.03	2.923	5.46	1.827	3.57	2.264
Oct.	−0.05	2.924	5.91	1.551	3.86	2.125
Nov.	−0.05	2.928	6.39	1.179	4.18	1.964
Dec.	−0.05	2.932	7.00	0.000	4.58	1.719

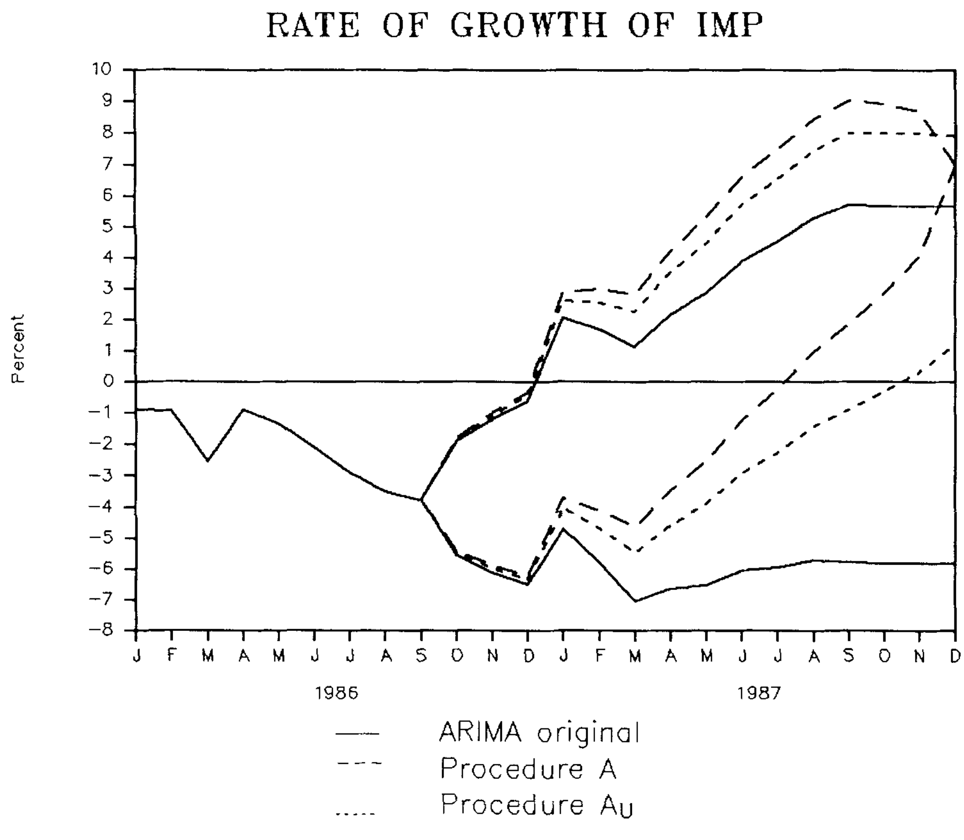

Figure 1. Ninety-five per cent confidence bands for the forecasts obtained with the different procedures.

In order to specify the 1×1 matrix U of procedure A_U we must solve inequality (48), that is,

$$[Y - \hat{R}_N^{\mathrm{IMP}}(15)]^2(\hat{\sigma}_a^2 C\hat{\Psi}\hat{\Psi}'C' + U)^{-1} < \chi_1^2(\alpha)$$

which yielded

$$U > 49.7025/\chi_1^2(\alpha) - 8.5978$$

Therefore for $\alpha = 0.05$ we must choose $U > 4.3456$.

The original ARIMA forecasts, together with the conditional ones obtained with procedures A and A_U (for which U was given the value 4.5, thus resulting in $\hat{K}_U = 3.795$) are shown in Table 1. In this table we also give the values of the forecast standard errors.

The pattern of the confidence bands derived from the original and the conditional ARIMA forecasts can be observed in Figure 1, where we can appreciate the increase of precision due to the additional information and the effect of considering the additional information as a given fact or conjecture.

Another illustration

A slightly more complicated situation arises when we have more than one restriction. Let us consider, for instance, that not only the rate of growth of IMP has to reach 7% in December 1987 but also that the average of the rates during that year has to be 3%. In this case we have $H = 15$, $m = 2$,

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & 1/12 & \dots & 1/12 & 1/12 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

The calculated statistics in this situation became $\hat{K}_1 = 5.781$ and $\hat{K}_2 = 2.891$ which, when compared with $\chi_2^2\ (0.1) = 4.61$ and $F_{2,127}(0.1) = 2.35$, lead us to reject the hypothesis of

Table 2. Forecasts for the rate of growth of IMP: $\mathbf{Y} = (7, 3)'$

	ARIMA		Procedure A		Procedure $A_U(U = 3.51)$	
Month	Forecasts	Std. error	Forecasts	Std. error	Forecasts	Std. error
Oct. 1986	−3.72	0.931	−3.59	0.850	−3.40	0.870
Nov.	−3.66	1.251	−3.42	0.996	−3.03	1.061
Dec.	−3.55	1.504	−3.25	0.915	−2.63	1.091
Jan. 1987	−1.31	1.721	−0.38	0.953	0.23	1.118
Feb.	−2.04	1.913	−0.54	0.931	0.06	1.093
March	−2.94	2.087	−0.86	0.890	−0.31	1.042
April	−2.23	2.248	0.42	0.848	0.91	0.979
May	−1.82	2.398	1.41	0.818	1.80	0.918
June	−1.07	2.539	2.73	0.808	3.01	0.873
July	−0.71	2.673	3.66	0.821	3.80	0.855
Aug.	−0.19	2.801	4.75	0.850	4.73	0.871
Sep.	−0.03	2.923	5.48	0.883	5.28	0.919
Oct.	−0.05	2.924	5.93	1.056	5.36	1.184
Nov.	−0.05	2.928	6.40	0.995	5.45	1.331
Dec.	−0.05	2.932	7.00	0.000	5.56	1.332
Average 1987	1.041	–	3.000	–	2.990	–

consistency between conjecture and model at the $\alpha = 0.1$ significance level. Therefore an application of procedure A_U was required.

Specifying the 2×2 matrix U is not as simple now as it was in the previous illustration. Nevertheless, by trial and error and assuming the form $U = \sigma_u^2 I$, we found that a value $\sigma_u^2 = 3.5$ was large enough to satisfy criterion (48), since for this particular value we obtained

$$\hat{K}_U = 4.606 < \chi_2^2(0.1)$$

Therefore $U = 3.5$ was considered appropriate. The results of applying procedures A and A_U are summarized in Table 2.

CONCLUSIONS

Two statistical procedures for obtaining conditional forecasts based on historical data and some linear condition on the forecasts have been presented. Both were derived from the same ARIMA model constructed solely on the basis of the observed data. The way in which the linear condition applies (1) as a given fact or (2) as conjecture produces different results, whose appropriateness for the data can be judged by means of test statistics derived here.

It is worth noting that the generalization of the fifth section of this paper provides a solution to the problem of combining forecasts from two different models, one of which is a univariate ARIMA model. This can be easily seen by choosing C as the identity matrix of order H in expression (40).

Finally, the computations involved should not pose any difficulty for applying the procedures in practice, since the most difficult part is to obtain the numerical value of the $m \times m$ matrix $(C\psi\psi' C' + U/\sigma_a^2)^{-1}$, where m is typically a small number. On the other hand, the illustrative examples, both theoretical and empirical, demonstrate the potential usefulness of the procedures, which can be valuable tools for an applied time series analyst or an econometrician.

APPENDIX: SOLUTION TO THE GENERALIZED PROBLEM

Here we show the basic steps followed by Pérez-Porrúa in his derivation of expressions (42) and (43). This method is essentially the same as that used by Doan *et al.* (1983), but the closed expressions obtained here allow us to understand it better. We have changed the notation used by Pérez-Porrúa to make it consistent with the notation used throughout this paper.

The main objective is to employ the iterated expectation formula

$$E(\mathbf{Z}_F \mid \mathbf{Z}_0, \mathbf{Y}) = E(\mathbf{Z}_F \mid \mathbf{Z}_0) + E[(\mathbf{Z}_F - E(\mathbf{Z}_F \mid \mathbf{Z}_0)) \mid \mathbf{Y} - E(\mathbf{Y} \mid \mathbf{Z}_0)]$$

which follows from the Law of Iterated Projections. Then from equations (40) and (4) we have

$$\begin{aligned} \mathbf{Y} - E(\mathbf{Y} \mid \mathbf{Z}_0) &= (C\mathbf{Z}_F + \mathbf{u}) - CE(\mathbf{Z}_F \mid \mathbf{Z}_0) \\ &= C\psi \mathbf{a}_F + \mathbf{u} \end{aligned}$$

so that

$$E[(\mathbf{Z}_F - E(\mathbf{Z}_F \mid \mathbf{Z}_0)) \mid \mathbf{Y} - E(\mathbf{Y} \mid \mathbf{Z}_0)] = E(\psi \mathbf{a}_f \mid C\psi \mathbf{a}_F + \mathbf{u})$$

Now, by the normality assumption, the last expectation is linear, so that

$$E(\psi \mathbf{a}_F \mid C\psi \mathbf{a}_F + \mathbf{u}) = A_U(C\psi \mathbf{a}_F + \mathbf{u})$$

where A_U is an $H \times m$ matrix, chosen in such a way to satisfy the orthogonality condition

$$E\{(C\psi\mathbf{a}_F + \mathbf{u})[\mathbf{a}_F'\psi' - (\mathbf{a}_F\psi'C' + \mathbf{u}')\hat{A}_U']\} = 0$$

which leads to the solution

$$\hat{A}_U = \psi\psi'C'(C\psi\psi'C' + U/\sigma_a^2)^{-1}$$

as was indicated in equation (43).

ACKNOWLEDGEMENTS

The author acknowledges the benefit of discussions on the topic of this paper with J. M. Pérez-Porrúa and J. Romo. He also thanks S. Portela for helping him with some of the computations.

REFERENCES

Box, G. E. P. and Jenkins, G. M., *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day, 1970.

Cholette, P. A., 'Prior information and ARIMA forecasting', *Journal of Forecasting*, **1** (1982), 375–83.

Doan, T., Litterman, R. and Sims, Ch., 'Forecasting and conditional projection using realistic prior distributions', Working Paper 243, Research Department, Federal Reserve Bank of Minneapolis, 1983.

Pérez-Porrúa, J. M., 'Pronósticos ARIMA que incorporan información externa', Unpublished manuscript, Dirección de Investigación Económica, Banco de México, 1984.

Vera, G. and Guerrero, V. M., 'Estimation of seasonal factors using both traditional methods and Box–Jenkins techniques', in Anderson, O. D. and Perryman, M. R. (eds), *Time Series Analysis*, Amsterdam: North-Holland, 1981.

Author's Biography:

Víctor M. Guerrero is a statistician at the Banco de México and part-time lecturer at the Instituto Politécnico Nacional. He has a PhD in statistics from the University of Wisconsin-Madison. His present research interests are mainly in the area of time series analysis and forecasting. He has published articles in *Biometrika*, *Insurance*: *Mathematics and Economics*, *Letters in Probability and Statistics* and *Economics Letters* and has some other publications in Spanish, including a textbook.

Author's address:

Víctor M. Guerrero, Dirección de Investigación Económica, Banco de México, Av. Juárez 90, México 06040 DF, México.