Statistical and methodological issues in short and medium term forecasting[1]

G.STORTI[a] & C. VITALE[a]

[a]*Università degli Studi di Salerno, Dipartimento di Scienze Economiche, Via Ponte Don Melillo, 84084 Fisciano, Italy-email:vitale@unisa.it*

ABSTRACT

This paper reviews some of the most important practical and theoretical issues arising in the implementation of a stochastic forecasting model for an observed time series of data. Different general approaches to calculating forecasts are described and an overview of some the most recent developments in the field of time series analysis and forecasting is given. Particular emphasis is placed on the importance of model selection biases and on techniques which allow to limit the effects of model uncertainty.

[1]Paper supported by MURST 1998: Modelli Statistici per l'Analisi delle Serie Temporali. The work is joint responsibility of the authors; C. Vitale wrote sections 1,6; G. Storti wrote sections 2,3,4 and 5

1 INTRODUCTION AND NOTATION

Let $\{\mathbf{Y}_t\} = \{Y_{t,1}, \ldots, Y_{t,m}\}$ be an m-dimensional observed time series, forecasting is the problem of predicting the future value of $\mathbf{Y}_{t+k}$, with $k > 1$ and integer, on the basis of the set of information available at time t, $\mathbf{D}^t = \{\mathbf{D}_1, \ldots, \mathbf{D}_t\}'$, where $\mathbf{D}_j$, $j = 1, \ldots, t$, is an s-dimensional row vector with s corresponding to the number of variables used to obtain information on $\{Y_t\}$. Hence, $\mathbf{D}^t$ will have dimension $(ts)\times 1$. We will denote this prediction by $\mathbf{y}(t,k)$ with $\mathbf{e_y}(t,k) = [\mathbf{Y}_{t+k} - \mathbf{y}(t,k)]$ being the associated prediction error. If, conditional on $\mathbf{D}^t$, $E[\mathbf{e_y}(t,k)] = \mathbf{0}$, the predictor $\mathbf{y}(t,k)$ will said to be unbiased. An usual yardstick for comparing different forecasts is given by the *Mean Squared Prediction Error* (PMSE)

$$PMSE(\mathbf{y},k) = E\{[e_{\mathbf{y}}(t,k)]^2\} \tag{1}$$

If the predictor is unbiased, $PMSE(\mathbf{y},k) = var\{[\mathbf{e_y}(t,k)]\}$. In the univariate case ($m = 1$), the predictor which minimizes $PMSE(\mathbf{y},k)$ is the conditional expectation

$$\hat{y}(t,k) = E(Y_{t+k}|\mathbf{D}^t) = G(\mathbf{D}^t). \tag{2}$$

In the more general multivariate case ($m\neq 1$) the optimality of the predictor $\hat{\mathbf{y}}(t,k)$ must be properly reinterpreted in the sense that the PMSE matrix of any other predictor can be expressed as the PMSE matrix of the conditional expectation $\hat{\mathbf{y}}(t,k)$ plus a matrix which is positive semidefinite.

The functional form of $\mathbf{G}(.)$ depends on the shape of the conditional density $p(\mathbf{Y}_{t+k}|\mathbf{D}^t)$, also called the *predictive density*. In a strict sense, forecasting can be then redefined as the problem of calculating the predictive density $p(\mathbf{Y}_{t+k}|\mathbf{D}^t)$. At this stage, two remarks have to be made. First, in general the shape of the density will not be the same for different values of the lead time k. Second, given a lead time k, the functional form of the predictive density will not necessarily remain unchanged as t varies.

In general the shape of $\mathbf{G}(.)$ is not completely known and has to be estimated from the observed data. A common approximation is to assume that the function $\mathbf{G}$ is linear giving the predictor

$$\hat{y}_l(t,k) = \mathbf{G}(t,k)\mathbf{D^t} \tag{3}$$

where $\mathbf{G}(t,k)$ has dimension $(m\times ts)$ and is given by

$$\mathbf{G}(t,k) = E[\mathbf{Y_{t+k}}(\mathbf{D^t})']\{E[\mathbf{D^t}(\mathbf{D^t})']\}^{-1}. \tag{4}$$

In practical applications the calculations can be simplified by formulating general assumptions on the dependence structure of the process, such as Markovianity, which reduce the dimension of the prediction coefficient $\mathbf{G}(t,k)$ by causing some elements of $\mathbf{G}(t,k)$ to be equal to 0.
As shown in Catlin (1989), a recursive expression for the *Linear Least Squares Predictor* (LLSP) is obtained as

$$\begin{aligned} E[\mathbf{Y}_{t+k}|\mathbf{D}^t] &= E[Y_{t+k}|\mathbf{D}^{t-1},\mathbf{D}_t] = E[Y_{t+k}|\mathbf{D}^{t-1}] \\ &+ cov(\mathbf{Y_{t+k}},\mathbf{D_t}|\mathbf{D^{t-1}})var(\mathbf{D_t}|\mathbf{D^{t-1}})^{-1}[\mathbf{D_t} - E(\mathbf{D_t}|\mathbf{D^{t-1}})]. \end{aligned}$$

In the case in which the joint distribution of $\mathbf{Y}_{t+k}$ and $\mathbf{D}^t$ is multivariate normal, for all t, the LLSP (3) is equal to the conditional expectation of $\mathbf{Y}_{t+k}$ given $\mathbf{D}^t$.
Aim of this paper is to analyse some of the main problems which arise in the construction of a stochastic model for forecasting purposes. In section 2 a description of the possible general approaches to the problem of estimating the shape of $\mathbf{G}(.)$ is given while some specific issues related to forecasting in the framework of state-space models are dealt with in section 3. Section 4 is concerned with the calculation of prediction intervals as a measure of uncertainty for the obtained forecasts, conditional on the assumption that the data generating mechanism is known. The effect of misspecification errors at the model selection stage is the subject of section 5. Section 6 concludes.

2 APPROACHES TO FORECASTING

In this section we will examine the different modeling approaches which can be adopted in order to estimate the function $G(.)$

1. Parametric methods
   (a) Linear models (ARMA)
   (b) Non-linear models
2. Non-parametric methods

In the first case, the functional form of $\mathbf{G}(.)$ is assumed known but its parameters have to be estimated while, in the second case, we do not formulate any assumptions on the functional form of $\mathbf{G}(.)$ whose shape has to be estimated numerically from the data.

2.1 Linear ARMA models

Linear Autoregressive Moving Average (ARMA) models (Box and Jenkins, 1970) have been successfully applied in many situations for modeling economic

and environmental phenomena. Several applications to hydrological problems have also been proposed in the literature (e.g. Claps *et al.*, 1993). The book by Piccolo and Vitale (1981) provides an introductory reading to the topic while a more detailed account is given by Brockwell and Davis (1987).

Consider a linear ARMA(*p*,*q*) model for an univariate series $\{Y_t, t = 1, \ldots, T\}$ with $E[Y_t] = 0$

$$\phi_p(B)Y_t = \theta_q(B)Z_t$$

where $Z_t \sim N(0, \sigma_z^2)$ is a serially uncorrelated *white noise* process and

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \\ \theta_p(B) &= 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \end{aligned}$$

are polynomials in the *backward operator* B having no common zeros. The model is said to be *invertible* if $\theta(B) \neq 0$ for $|B| < 1$. An invertible ARMA model always admits the AR(∞) representation

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \ldots + Z_t \tag{5}$$

This can be rearranged to give

$$Z_t = \pi_\infty(B)Y_t = Y_t - \pi_1 Y_{t-1} - \pi_2 Y_{t-2} - \ldots \tag{6}$$

where $\pi_\infty(B) = \phi_p(B)/\theta_q(B)$. In (6) current innovations are expressed as a linear function of the history of the observed process up to the infinite past, for some sequence of weights $\{\pi_i\}$ such that $\sum_{i=0}^{\infty} |\pi_i| < \infty$. Similarly, if $\phi(B) \neq 0$ for $|B| < 1$ the model is said to be *stationary* and the following MA(∞) representation holds

$$Y_t = \psi_\infty(B)Z_t \tag{7}$$

with $\psi_\infty(B) = \theta_p(B)/\phi_q(B)$, leading to the *linear filter*

$$Y_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}, \qquad \sum_{i=0}^{\infty} |\psi_i| < \infty \tag{8}$$

As we will see more closely in the next section, in general, for a non-linear process, this linear filter representation will be replaced by some non-linear function $g(\ldots, Z_{t-1}, Z_t)$.

When building forecasts from an ARMA model, invertibility and stationarity are crucial properties. While it is true that a given ARMA model has a unique

covariance structure, the converse is not true. However, although there exists a multiciplity of ARMA models possessing the same autocovariance function, there will be only *one stationary invertible* model which has a given autocovariance function. Furthermore, as shown in Box and Jenkins (1970), the invertibility condition ensures us that the dependence of the estimated innovations $\hat{z}_t$ on the estimated initial values $\{\hat{z}_0, \ldots, \hat{z}_q\}$ will tend to decrease rapidly as t increases.

Let us assume to know all the past history of the process

$$\mathbf{D}^t = \{\ldots, y_{-1}, y_0, y_1, \ldots, y_t\},$$

The k-steps ahead minimum PMSE predictor for an ARMA process is recursively generated from *known past* values y_{t-i} of the series and past values z_{t-i} which, because of the invertibility condition, can also be considered as known

$$\begin{aligned}\hat{y}_{t,k} &= E[\sum_{i=1}^{p} \phi_i Y_{t+k-i} + Z_{t+k} - \sum_{i=1}^{q} \theta_i Z_{t+k-i} | \mathbf{D}^t] \\ &= \sum_{i=1}^{p} \phi_i E[Y_{t+k-i}|\mathbf{D}^t] + E[Z_{t+k}|\mathbf{D}^t] - \sum_{i=1}^{q} \theta_i E[Z_{t+k-i}|\mathbf{D}^t]\end{aligned}$$

where

$$E[Y_{t+k-i}|\mathbf{D}^t] = \begin{cases} y_{t+k-i} & \text{for } i \geq k \\ \hat{y}_{t,k-i} & \text{for i} < k \end{cases}$$

$$E[Z_{t+k-i}|\mathbf{D}^t] = \begin{cases} z_{t+k-i} & \text{for } i \geq k \\ 0 & \text{for i} < k \end{cases}$$

From the MA(∞) representation in (8) it follows that the prediction error is given by

$$e_y(t,k) = Z_{t+k} + \psi_1 Z_{t+k-1} + \ldots + \psi_{k-1} Z_{t+1} \tag{9}$$

with prediction variance

$$var[e_y(t,k)] = \sigma_z^2[1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{k-1}^2] \tag{10}$$

which increases with the lead-time k. In practical applications, the parameters ψ_i are not known and have to be replaced by their estimates. In a multivariate setting an analogous result holds for Vector ARMA processes (Reinsel, 1997; Lütkepohl, 1993).

2.2 Non-linear models

In this paragraph we will examine some of the problems arising when we use non-linear models for forecasting purposes. Before moving to the analysis of some technical issues arising in forecasting from non-linear models, we will try to make clear what is respectively meant by a *linear* and a *non-linear* model.

2.2.1 Non-linear models: theoretical issues on non-linearity

It is well known that any stationary time series $\{Y_t\}$ with a purely continuous spectrum can be represented in the mean square convergent bilateral series

$$Y_t = \sum_{i=-\infty}^{\infty} g_i Z_{t-i} \tag{11}$$

where $\{Z_t\}$ is a sequence of uncorrelated random variables. As pointed out by Priestley (1980), this gives us a *linear representation* for $\{Y_t\}$ but does not constitute a model since the joint distribution of the $\{Z_t\}$ is, to a large extent, unspecified. Strictly, a *linear model* for $\{Y_t\}$ is a representation of the form (11) in which the Z_t are strictly *independent*. In particular, it can be shown (Nisio, 1960) that any stationary time series $\{Y_t, t = 0, \pm 1, \dots\}$ can be approximated, in a certain sense, by an *almost surely* convergent *polynomial representation* of the form

$$Y_t = \sum_{p} \sum_{i_1,\dots,i_p} a_{i_1,\dots,i_p} \prod_{l=1}^{p} Z_{t+i_l} \tag{12}$$

with $\{Z_t, t = 0, \pm 1, \dots\}$ being a sequence of i.i.d. random variables. If the summations over the i_j range from $-\infty$ to ∞, we say that the representation is *two-sided* while, if they range each over the interval (0,∞), the representation is then said to be *one-sided*. Furthermore, if $p = 1$, $\{Y_t\}$ is defined to be a *linear* time series while, for $p > 1$, $\{Y_t\}$ is called a *non-linear* time series.

Some special properties characterize Gaussian processes. First, it can be shown that any stationary Gaussian time series admits a two-sided linear representation. Second, any stationary Gaussian time series $\{Y_t\}$ admits a one-sided linear representation if it has a purely continuous spectrum whose density $f(.)$

a) is positive almost everywhere,

b) satisfies the Paley-Wiener condition $\int_{-\pi}^{\pi} \ln f(\omega) d\omega > -\infty$,

and Z_t is measurable with respect to the σ-algebra $\mathbb{B}_t$ generated by $Y_s, s \leq t$. For more general stationary time series the situation is not completely resolved. The interested reader may refer to Rosenblatt (1971) for more details.

2.2.2 Non-linear models: practical issues

The first consideration which has to be made, when trying to give an overview of the problems which arise when forecasting from non linear models, is that general solutions are not available but each different non-linear structure requires an *ad hoc* treatment. Nevertheless, some general difficulties are common to many of the non-linear models used by statisticians and practitioners

a) stochastic properties of the model are often unclear,

b) calculation of k-steps ahead forecasts is difficult or, in many cases, not possible at all,

c) it is not easy to verify the invertibility of the model.

A well known result for non linear models states that the prediction variance $var(e_y(n,k))$ is monotone in the lead time k. This will not in general be true for non-linear models. Some theoretical results and a discussion are given by Tong (1990).

A rather general model structure encompassing a wide range of non-linear processes is defined as

$$Y_t = m(\mathbf{D}^{t-j}, \boldsymbol{\alpha_1}) + \sigma(\mathbf{D}^{t-j}, \boldsymbol{\alpha_2}) Z_t \tag{13}$$

where $\{Z_t\}$ is a series of zero mean i.i.d. random variables and $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_2}$ are vectors of unknown parameters. The functions $m(.)$ and $\sigma(.)$ are known with $E(Y_t|\mathbf{D^{t-j}}) = m(\mathbf{D^{t-j}}, \boldsymbol{\alpha_1})$ and $var(Y_t|\mathbf{D^{t-j}}) = \sigma^2(\mathbf{D^{t-j}}, \boldsymbol{\alpha_2}) var(Z_t)$. An usual convention is to assume $var(Z_t) = 1$. Several models can be obtained as a special case of (13) for different choices of the $m(.)$ and $\sigma(.)$ functions.

Directly from (13), assuming that $\sigma(.)$ is constant and equal to unity and that the conditional mean is a function only of past realizations of the process, we obtain the class of *non-linear autoregressive* (NLAR) models with additive noise

$$Y_t = m(Y_{t-1}, \ldots, Y_{t-p}) + Z_t \tag{14}$$

with $m : \mathbf{\Re^p \rightarrow \Re}$. These models are invertible by definition. Recursive formulae for the calculation of the predictive density and its expectation are given by Tong(1990). Different choices of $m(.)$ yield some well known non-linear models.

In Threshold AutoRegressive (TAR) models the conditional mean function $m(.)$ is piecewise linear and the conditional variance $\sigma^2(.)$ is set equal to the identity function. Let K_t be an appropriately chosen *threshold variable* whose lagged

values influence the actual state of the system and $\{R_1, \ldots, R_l\}$ a partition of the real line $\Re$. A TAR$(l, p_1, \ldots, p_l)$ model for $\{Y_t\}$ can be written as

$$Y_t = \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} Y_{t-i} + Z_t^{(j)} \tag{15}$$

conditional on $K_{t-d} \in R_j$, with $d > 0$, where $\{Z_t^{(j)}\}$ are heterogeneous white noise sequences with zero mean and finite variances, each being independent of $\{Y_t\}$. The model structure in (15) can be generalized in order to allow for exogenous inputs $\{W_t\}$ by considering the class of Open Loop Threshold AutoRegressive (TARSO) models (Tong & Lim, 1980)

$$Y_t = \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} Y_{t-i} + \sum_{i=1}^{r_j} \beta_i^{(j)} W_{t-i} + Z_t^{(j)} \tag{16}$$

for $K_{t-d} \in R_j$. Successful applications of threshold models in hydrology can be found in Tong & Lim (1980) and Tsay (1998). In a recent paper, Amendola & Storti (1999) suggest using a TARSO for analysing the rainfall-flow process. The inherent non-linearity of this relationship is typically due to the effect of variations in the catchment's moisture conditions. The modeling procedure proposed in the paper is illustrated by means of an application to data from a small basin in Southern Italy, Sarno river at San Valentino Torio (SA, Italy). The state-dependent nature of the process is captured by using the Antecedent Precipitation Index (API) as a threshold variable (Fig. 1). Also, their model allows for different autoregressive orders within each regime in order to reproduce the changing behaviour of the system as the API crosses various thresholds. Although the catchment analysed in the paper has a reduced extension (42 km^2), the TARSO model gives a satisfactory forecasting performance (Fig. 2). In general, threshold autoregressive models are particularly useful for reproducing asymmetries in the relationship between past information and the conditional mean of a time series. Forecasting from models of this class does not create particular problems provided that the lead time k does not exceed the value of the delay d, $k \leq d$. If $k > d$ we do not have sufficient information to identify the regime in which the process will be situated at time $t + k$, unless we build a forecasting model for the threshold variable K_{t-d}. It is however worth noting how, using predicted values of K_{t-d}, we are going to introduce into the model an additional source of uncertainty which can be not easily controllable.

An alternative to threshold models is given by Switching Regime models (Hamilton, 1989) in which regime changes are regulated by the transition probability of an hidden Markov chain and not by lagged values of a threshold variable.

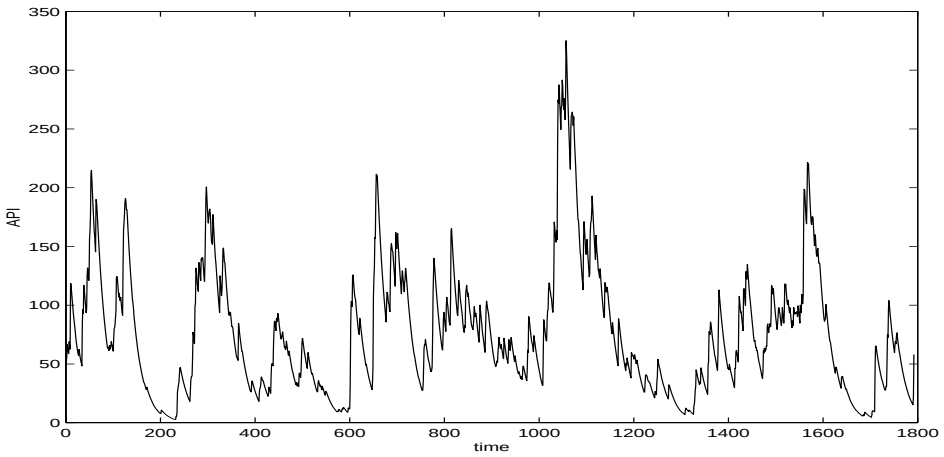

Figure 1: Antecedent Precipitation Index (mm/d) from 1.1.74 to 8.3.78; River Sarno at San Valentino Torio (SA, Italy).

The general class of NLAR models described in (14) enjoys some attractive probabilistic properties. Namely, assume that

a) the initial values $Y_i = y_i$ are known for $i = -p+1, \ldots, -1, 0$,

b) $m(.)$ is bounded on compact sets,

c) $m(\mathbf{y}) = \boldsymbol{a}^{\boldsymbol{T}}\mathbf{y} + o(\|\mathbf{y}\|)$ as $\|\mathbf{y}\| \to \infty$ and provided that $\boldsymbol{a}^{\boldsymbol{T}}\mathbf{y}$ is stable,

d) the density of Z_t is positive everywhere and $E(|Z_t|) < \infty$.

These are sufficient conditions for the existence of a stationary distribution for $\mathbf{Y}^0 = \{Y_0, Y_{-1}, \ldots, Y_{-p+1}\}$ such that if Y_t is started in this distribution, it is strictly stationary. The above conditions also imply that the model is geometrically ergodic and strongly α-mixing. The mixing property, in turn, allows to apply central limit theorem for strongly mixing processes (Tjøstheim, 1994).

A different kind of non linearity arises from the specification of a parametric model for the conditional variance. In Generalized AutoRegressive Conditionally Heteroskedastic (GARCH) models (Bollerslev, 1986) the conditional mean is constant and equal to zero while the conditional variance is a linear function of past squared observations and past conditional variances

$$\sigma^2(\mathbf{D^{t-1}}) = a_0 + \sum_{i=1}^{p} a_i y_{t-i}^2 + \sum_{i=1}^{q} b_i \sigma^2(\mathbf{D^{t-i-1}}) \qquad (17)$$

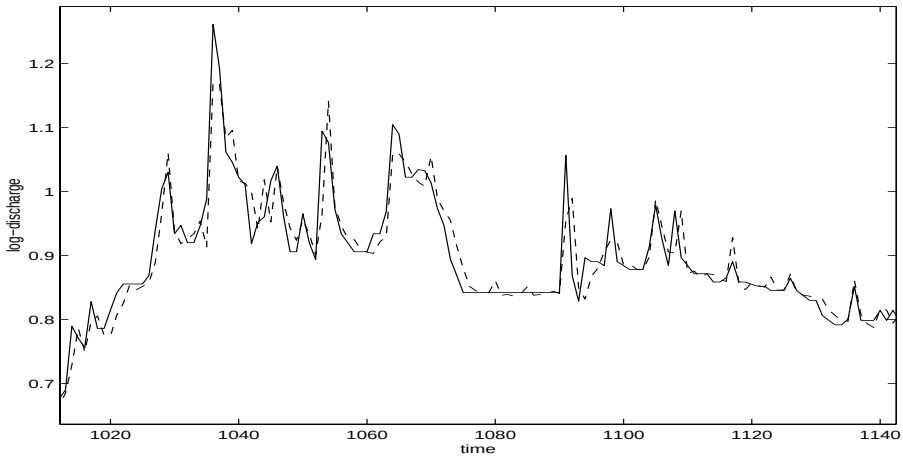

Figure 2: 3-hours-ahead forecast (dashed line) and observed data (solid line) for the high flow period starting October 10, 1976 and ending February 14, 1977 (m^3/s); River Sarno at San Valentino Torio (SA, Italy).

with $a_0 > 0$, $a_i \geq 0$ (for i=1,. . . ,p), $b_j \geq 0$ (for j=1,. . . ,q), $p > 0$ and $q \geq 0$. Models of this kind are often specified for the residuals from a time series model for the conditional mean in order to capture the changing conditional variance of the process.

It is possible to define a third class of models, the so called *mixed* non-linear models, that are characterized by the presence of non-linear effects in both the conditional mean and variance functions. As an example consider the class of Self Exciting Threshold Autoregressive Models with AutoRegressive Conditional Heteroskedasticity (SETAR-ARCH) proposed by Li & Lam (1995). Non-linearities in both $m(.)$ and $\sigma(.)$ also characterize the well known family of BiLinear (BL) models (Granger & Andersen, 1978). A BL model has the general form

$$Y_t = \phi_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=0}^{q} \theta_i Z_{t-i} + \sum_{j=1}^{r} \sum_{k=1}^{s} \delta_{jk} Y_{t-j} Z_{t-k} \qquad (18)$$

where the i.i.d. errors Z_t are usually, but not always, assumed to be zero mean, the ϕ, θ and δ are real unknown constants with $\theta_0 = 1$. The model can be considered as an extension of an ARMA(p,q) structure in which the non-linearity stems from the interaction terms $\delta_{jk} Y_{t-j} Z_{t-k}$, for j=1,. . . ,r and k=1,. . . ,s. A difficulty commonly arising with BL models is that it is often not possible calculate explicit expressions for the conditional mean and variance specifications.

Evaluation of point predictions for BL models is quite straightforward provided

that the process is invertible. Unfortunately sufficient conditions for invertibility are available in isolated cases only and it is not easy to verify them in practice. If the process is not invertible we are going to incur problems analogous to those described for ARMA models.

In general, it is possible to specify conditions analogous to $(a) - (d)$ also for models characterized by a time varying conditional variance (Masry & Tjøstheim, 1995). Results present in the literature refer mainly to ARCH-related models (Diebolt & Guegan, 1991; Bouguerol & Picard, 1992; Rudolph, 1998).

2.3 Non-parametric approach

Consider again the general model structure (13). In many situations it is not possible to achieve a reliable identification of the parametric specifications for $m(.)$ and $\sigma(.)$. This can happen due to several reasons, such as lack of background knowledge on the phenomenon or reduced availability of data. However, in cases in which $m(.)$ and $\sigma(.)$ are unknown, we can still consider estimating them by using non-parametric procedures.

In the simple case of a constant conditional variance, $m(\mathbf{D}^{t-j})$ can be estimated by direct application of standard non-parametric techniques such as Kernel and local polynomial regression methods (for a review see Härdle *et al.*, 1997). In this context calculation of 1-step ahead forecasts is relatively straightforward. Some modifications are required in order to use non-parametric techniques for the purpose of obtaining multi-step ahead predictions. Referring to the case of a NLAR model of order one with additive errors, Härdle (1990) and Härdle and Vieu (1992) proposed using the ordinary Nadaraya-Watson Kernel estimator to estimate $E[Y_{t+k}|\mathbf{Y}^t]$ directly. The estimator they use is

$$\hat{m}_{h,k}(y) = \frac{\sum_{t=1}^{T-k} K\{(y - Y_t)/h\} Y_{t+k}}{\sum_{t=1}^{T-k} K\{(y - Y_t)/h\}} \tag{19}$$

where $K(.)$ is a kernel function with bounded support and h is the bandwidth. For more details on Kernel smoothing see Härdle (1990). Chen and Hafner (1995) propose an alternative procedure which makes use of the information contained in intermediate variables $\{Y_{t+1}, \ldots, Y_{t+k-1}\}$ in order to estimate $E[Y_{t+k}|\mathbf{Y}^t]$. Their proposal is based on a multistage Kernel smoother which is shown to have a smaller PMSE than (19).

Various non parametric approaches also exist for joint estimation of $m(.)$ and $\sigma(.)$ in the more general case of a non-constant conditional variance. A very widely diffused approach is based on the direct estimator

$$\hat{\sigma}_d^2(y) = \hat{v}(y) - \{\hat{m}(y)\}^2 \tag{20}$$

where $\hat{m}(y)$ and $\hat{v}(y)$ are respectively a regression estimator (e.g. a Kernel) for $m(y)$ and $v(y) = E[Y^2|\mathbf{D^{t-1}}]$ (Härdle and Tsybakov, 1997). The main problem with this approach is that the estimated conditional variance is not always non-negative. Furthermore it can produce a substantially large bias. In order to overcome these difficulties, Fan & Yao (1998) suggest an alternative procedure based on the application of local linear regression techniques to the squared residuals obtained from the model for the conditional mean. Some proposals for the application of non-parametric methods in order to estimate the full one-step ahead predictive density have also arisen in the literature (Robinson, 1983; Gallant e Tauchen, 1989).

Despite their flexibility, non-parametric procedures also suffer from some drawbacks. First, the asymptotic mean squared error of non-parametric estimates is larger than that of parametric ones and tends to increase with the dimension m of the model. This leads to consider a second more serious problem, the so called *curse of dimensionality*. As m increases, an extremely large sample size is required to get a sufficient number of points in the unit cube of the resulting m-dimensional space. Third, non-parametric estimates are less easily interpretable than those obtained from a parametric model in which a precise physical meaning is often associated to the estimated parameters. Furthermore, it is important to recognize that non-parametric procedures are not assumption free. As in the parametric case, for univariate models, we still need to identify the relevant lags of the response variable or, if the model is multivariate, the relevant explanatory variables to be included in the model. Furthermore the application of a non-parametric technique often requires assumptions about smoothing parameters and weight functions which can play a crucial role in determining the final performance of the estimator. If we undersmooth the observed process, in the limit our estimated curve will tend to reproduce the original data being perturbated by the error component. In the opposite case, if an excessive amount of smoothing is performed, some relevant information will be inevitably lost. One of the most important issues in the application of Kernel estimators is the selection of the bandwidth parameter h. To this purpose, several data driven procedures have been proposed in the literature like, for example, the *leave-one-out* cross validation criterion extended by Härdle and Vieu (1992) to the case of dependent observations. A related problem is the determination of the number of hidden neurons in a feed-forward neural-network which, as shown in Giordano & Perna (1999), can be considered as a smoothing parameter.

A hybrid approach, which allows to conjugate advantages of both the parametric and the non-parametric approach, is to specify a parametric functional form only for some of the regressors in the model while using non-parametric tech-

niques to estimate the dependence relationship from the other regressors. Some applications are illustrated in Engle *et al.*(1986) and Shumway *et al.*(1988).

3 FORECASTING IN STATE SPACE MODELS

The general non-linear state space model can be written as

$$\mathbf{Y}_t = \mathbf{C}_t(\mathbf{X}_t, \mathbf{r}_t) \tag{21}$$

$$\mathbf{X}_t = \mathbf{A}_t(\mathbf{X}_{t-1}, \mathbf{q}_t) \tag{22}$$

where $\mathbf{C}_t(.)$ and $\mathbf{A}_t(.)$ are (m$\times$1) and (n$\times$1) known vector functions, $\mathbf{X}_t$ is a (n$\times$1) *state vector* and $\{\mathbf{r}_t\}$ (m$\times$1) and $\{\mathbf{q}_t\}$ (n$\times$1) are white noise series of mutually independently distributed errors. Also the distribution functions of the error terms $\{\mathbf{r}_t\}$ and $\{\mathbf{q}_t\}$ are assumed to be known for all t but not necessarily Gaussian [2]. Forecasting can then be redefined as the problem of calculating the density $p(\mathbf{Y}_{t+k}|\mathbf{D}^t)$. The first step is to obtain the one step-ahead predictive density of the state

$$\begin{aligned} p(\mathbf{X}_{t+1}|\mathbf{D}^t) &= \int_{-\infty}^{+\infty} p(\mathbf{X}_{t+1}, \mathbf{X}_t|\mathbf{D}^t) d\mathbf{X}_t \\ &= \int_{-\infty}^{+\infty} p(\mathbf{X}_{t+1}|\mathbf{X}_t) p(\mathbf{X}_t|\mathbf{D}^t) d\mathbf{X}_t \end{aligned} \tag{23}$$

$$\tag{24}$$

which can then be recursively updated to give the k-steps ahead state predictive density

$$p(\mathbf{X}_{t+k}|\mathbf{D}^t) = \int_{-\infty}^{+\infty} p(\mathbf{X}_{t+k}|\mathbf{X}_{t+k-1}) p(\mathbf{X}_{t+k-1}|\mathbf{D}^t) d\mathbf{X}_{t+k-1}. \tag{25}$$

Finally, the observation predictive density $p(\mathbf{Y}_{t+k}|\mathbf{D}^t)$ can be calculated solving the integral

$$p(\mathbf{Y}_{t+k}|\mathbf{D}^t) = \int_{-\infty}^{+\infty} p(\mathbf{Y}_{t+k}|\mathbf{X}_{t+k}) p(\mathbf{X}_{t+k}|\mathbf{D}^t) d\mathbf{X}_{t+k} \tag{26}$$

where $p(\mathbf{X}_t|\mathbf{D}^t)$ is the *filtering density* obtained as

$$p(\mathbf{X}_t|\mathbf{D}^t) = \frac{p(\mathbf{Y}_t|\mathbf{X}_t) p(\mathbf{X}_t|\mathbf{D}^{t-1})}{\int_{-\infty}^{+\infty} p(\mathbf{Y}_t|\mathbf{X}_t) p(\mathbf{X}_t|\mathbf{D}^{t-1}) d\mathbf{X}_t}. \tag{27}$$

[2] The model structure in (21)-(22) can be easily generalized to allow for exogenous deterministic inputs.

The moments of the predictive density can be obtained by numerically or analytically solving an integral of the kind

$$\mathbf{g}_{t+k|t} = E[\mathbf{g}(\mathbf{Y}_{t+k})|\mathbf{D}^t] = \int_{-\infty}^{+\infty} \mathbf{g}(\mathbf{Y}_{t+k}) f(\mathbf{Y}_{t+k}|\mathbf{D}^t) d\mathbf{Y}_{t+k}$$

for a generic function $\mathbf{g}(.)$. Also the unknown parameters in $\boldsymbol{\theta}$ can be estimated by maximizing the likelihood function which can be written out as

$$\begin{aligned} L(\mathbf{Y}_T; \boldsymbol{\theta}) &= \prod_{t=1}^{T} p(\mathbf{Y}_t|\mathbf{D}^{t-1}) \\ &= \prod_{t=1}^{T} \int_{-\infty}^{+\infty} p(\mathbf{Y}_t|\mathbf{X}_t) p(\mathbf{X}_t|\mathbf{D}^{t-1}) d\mathbf{X}_t. \end{aligned} \tag{28}$$

In the case of a linear Gaussian state space model with additive errors, under some general conditions, the Kalman filter (Kalman, 1960) provides the optimal solution to the problem of estimating the predictive density $p(Y_{t+k}|\mathbf{D}^t)$. In the non-linear and/or non-Gaussian case we need to solve the recursions (24)-(27). Analytical solution of the above integrals is feasible only in isolated cases (see West & Harrison, 1997). More frequently it is necessary to resort to numerical procedures based on

(a) numerical integration algorithms, e.g. Kitagawa (1987) and Storti *et al.* (1998) for an application to rainfall forecasting

(b) simulation techniques such as Markov Chain Monte Carlo (MCMC) methods. See Carlin *et al.* (1992) and Mariano & Tanizaki (1998) for applications to state space modeling and Gilks *et al.*(1996) for a general introduction.

Due to their higher degree of flexibility with respect to numerical integration techniques, in the past few years research has focused mainly on simulation based techniques. Furthermore MCMC techniques are computationally more efficient than numerical integration. However, especially in high dimensional problems, these procedures can still be computationally intensive requiring long elaboration times. The increasing computing power of modern PCs is gradually removing obstacles of this nature.

Storti *et al.*(1998) have applied non-linear state space techniques to short term rainfall forecasting. They suggest using a two-stage modeling procedure based on the definition of a series of disjoint *reference temporal windows* (RTW) in which the observations are assumed to be i.i.d.

$$\{Y_1, \ldots, Y_k\}, \{Y_{k+1}, \ldots, Y_{2k}\}, \ldots, \{Y_{T-k+1}, \ldots, Y_T\},$$

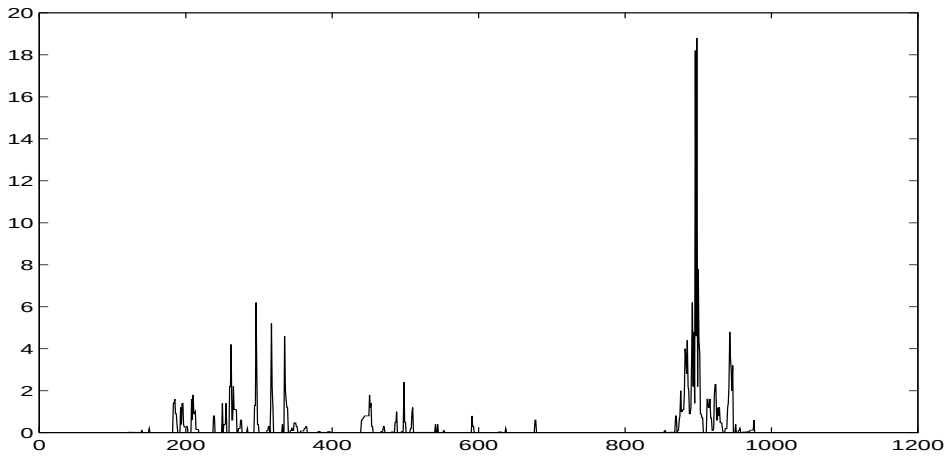

Figure 3: Total rainfall depth. Observations are taken at intervals of 20 minutes (Laverna, 24.10-6-11.1966).

with $\{Y_t, t = 1, \ldots, T\}$ being a time series of univariate rainfall observations taken at intervals of 20 minutes from 24.10.1966 to 6.11.1966 at Laverna, Italy, in the Arno Basin (Fig. 3).

At the first step, the number of occurrences observed in each RTW is used to estimate the probability of occurrence of rainy spells with intensity exceeding a given threshold within the next RTW. This is accomplished by applying NI techniques in order to estimate a non-linear state space model for the probability of occurence p_{t+1} (Fig. 4). Let $\hat{p}(N_{t+1}|Y^t)$ be the estimated probability mass function of the number of occurrences within the RTW (N_{t+1}), given past information Y^t. At the second step, an estimate of the predictive density function of the total rainfall depth $p(J_{t+1}|Y^t)$ is obtained solving the integral

$$\hat{p}(J_{t+1}|Y^t) = \int \hat{p}(J_{t+1}|N_{t+1})\hat{p}(N_{t+1}|Y^t)dN_{t+1}, \tag{29}$$

The predicted and filtered values of the total rainfall depth in each RTW have been reported in (Fig. 5).

4 PREDICTION INTERVALS

Point forecasts can be misleading if an adequate measure of uncertainty is not provided together with the calculated forecast. A measure of predictive uncertainty is often given in the form of a prediction interval for the point forecast.

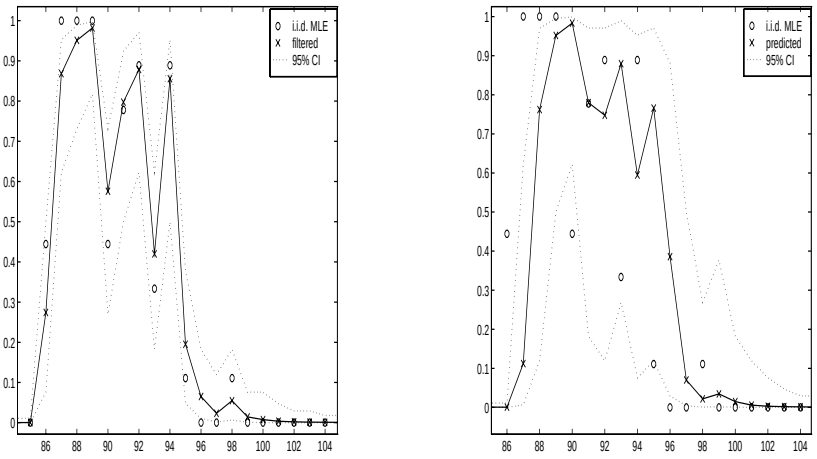

Figure 4: Filtered and predicted probabilities of occurrence for k=9 corresponding to a RTW of 3 hours (from 8.00 pm of 1.11.1966 to 4.00 am of 4.11.1966)

Also prediction intervals can be used to compare forecasts from different methods more thoroughly. Formally, a *prediction interval* (PI) can be defined as an interval forecast associated with a specified probability. A PI is associated with an estimate of an unknown future value that can be regarded as a random variable at the time the forecast is made. In order to make correct inferences on this random variable we need to estimate its probability density function conditional on current information $\mathbf{D}^t$. It follows that, strictly speaking, to set a PI for y_{t+k} we need to know the distribution of the forecast error $e_y(t,k)$ conditional on $\mathbf{D}^t$. In this section a description of some general approaches to the calculation of PIs is given. A distinction is made between approaches based on the application of theoretical formulas and empirical techniques based on simulation or resampling. A comprehensive review is given in Chatfield (1993).

A common practice is to assume that the forecast is unbiased with PMSE given by $E[e_y(t,k)^2] = var[e_y(t,k)]$ and that the forecast error is normally distributed. Under these assumptions, a 100(1-α) % PI for y_{t+k} is given by

$$y(t,k) \pm z_{\alpha/2} \sqrt{var\{e_y(t,k)\}} \tag{30}$$

where $z_{\alpha/2}$ denotes the appropriate percentage point of a standard normal distribution. When $var[e_y(t,k)]$ has to be estimated (as it happens in common applications) some authors (e.g. Harvey, 1989) suggest replacing the standardized normal distribution percentile $z_{\alpha/2}$ by the same order percentile $t_{\alpha/2}$ of a t distribution with an appropriate number of degrees of freedom. However this makes

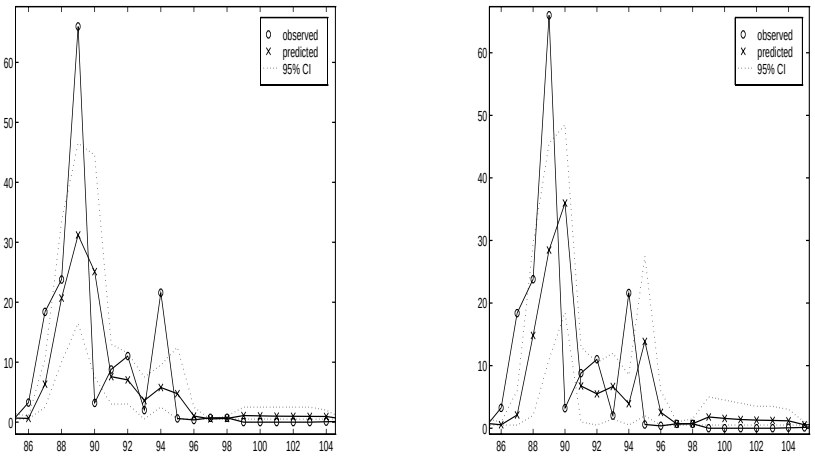

Figure 5: Filtered values and 1-*step ahead* forecasts of the total rainfall depth for k=9 corresponding to a RTW of 3 hours (from 8.00 pm of 1.11.1966 to 4.00 am of 4.11.1966)

little difference except for very short series (e.g. $T < 20$). For many models, the normality assumption of 1-step ahead forecast errors will be only asymptotically verified. Differently, for moderate sample sizes, the 1-step ahead predictive density will not in general be normal when model parameters have to be estimated from the same data used to calculate predictions (e.g. see Phillips, 1979). In these cases, before resorting to a normal approximation, it would be advisable to check the rate of convergence of the asymptotic approximation or at least analyze the empirical distribution of past forecast errors in order to quantify the deviation from normality.

When the normal approximation fails and *ad hoc* theoretical formulas are not available, the empirical distribution of errors can be used to estimate PIs. One possible way of exploiting the observed distribution of errors to set PIs is to use a *Bootstrap* procedure (for an introduction see Efron & Tibshirani, 1993). This class of resampling procedures involves simulating several pseudo-random replicates from the empirical distribution of past fitted errors in order to approximate the theoretical distribution of innovations. The procedure has the advantage of being free from distributional assumptions and easy to implement on a PC but can involve a considerable computational burden. In the past few years there has been a considerable number of papers proposing applications of Bootstrap procedures in the field of time series analysis. Excellent reviews are given by and Shao & Tu (1995) and Davison & Hinkley (1997). Among the others, Thombs

& Schucany (1990) illustrated an application to forecasting from AR processes, Kunsch (1989) and Politis & Romano (1994) explored the possibility of extending Bootstrap techniques to non-linear models and, more recently, La Rocca & Vitale (1999) have analysed the properties of Bootstrap estimates for BL models parameters.

An alternative approach consists in simulating several replicates of random innovations from a given probability time series model, instead of using the empirical distribution of past within-sample prediction errors. The set of replicates so obtained are sometimes called *pseudodata*. The main drawback of this approach is that it still relies on distributional assumptions and, as the Bootstrap method, is computationally intensive.

In the context of state space models the Bayesian forecasting procedures which have been illustrated in section (3) offer a natural approach to the problem of setting prediction intervals for y_{t+k}. The availability of the full predictive density $p(Y_{t+k}|\mathbf{D}^t)$ allows to set prediction bounds based on the percentiles of the estimated density with no need to resort to approximate distributional assumptions such as normality.

5 SOURCES OF UNCERTAINTY IN FORECASTING: REDUCING THE EFFECT OF MODEL UNCERTAINTY

In general when forecasts are calculated on the basis of a fitted statistical model we are implicitly assuming that

a) the identified model reflects the true data generating process,

b) the model parameters are exactly known,

c) the data are observed exactly.

Overall, the effects of parameter driven uncertainty seem likely to be smaller than those due to model uncertainty and to the presence of unexplained random variation in observed variables which can arise due to different reasons (uncorrectly measured variables, unknown initial values, outliers, exogenous variables in multivariate models have to be forecast). Statistical theory provides us tools for the treatment of outliers and for dealing with unknown initial values and measurement errors in the data. Measurement errors, for example, fit naturally in the context of state space models and, by means of slight modifications to the usual filtering algorithms, unknown initial conditions can be easily handled.

More severe consequences are likely to be produced by model misspecification errors. Although model uncertainty plays a crucial role in the forecasting process,

the effect of model misspecification can be hardly measured, even because we actually do not know the *true* model (if there is one). However, different solutions have been proposed in the statistical literature in order to limit the effects of model misspecification errors.

5.1 Mixture models

One possible approach for reducing the dependence from having to assume that there is a single *true* model, is to use a model which is actually a mixture of various different models with a prior probability weight associated to each of them. The observed data are then used to recursively update the weights. At each step, the Bayes rule is applied for calculating the correspondent posterior probabilities. Several applications of this approach in a state space setting have been illustrated by West & Harrison (1997). The class of models they propose, called *mixture models*, can be considered as a special case of the more general and well known technique of Bayesian *model averaging*. A recent review and methodology discussion for Bayesian model averaging is given by Draper (1995). Consider the linear state space model, or *Dynamic Linear Model* (DLM) in the terminology of West & Harrison (1997),

$$\mathbf{Y}_t = \mathbf{C}_t\mathbf{X_t} + \mathbf{r}_t \tag{31}$$

$$\mathbf{X_t} = \mathbf{A}_t\mathbf{X_{t-1}} + \mathbf{q}_t \tag{32}$$

where $\mathbf{q}_t \sim WN(\mathbf{0}, \mathbf{Q}_t)$, $\mathbf{r}_t \sim WN(\mathbf{0}, \mathbf{R}_t)$, $E(\mathbf{r_t}, \mathbf{q_s'}) = \mathbf{0}$, $\forall t, s$, $E(\mathbf{r_t}, \mathbf{X_0'}) = \mathbf{0}$, $E(\mathbf{q_t}, \mathbf{X_0'}) = \mathbf{0}$. Also assume that $(\mathbf{X}_0|\mathbf{Y_0}) \sim N(\mathbf{x_0}, \mathbf{P}_0)$. Given appropriately defined initial conditions, the model is uniquely identified by the quadruple

$$M_t \quad : \quad \{\mathbf{C}, \mathbf{A}, \mathbf{R}, \mathbf{Q}\}_t \tag{33}$$

Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ be a vector containing all the defining parameters of the model that are possibly subject to uncertainty. The notation $M_t(\boldsymbol{\theta})$ indicates that the model identified by the quadruple in (33) is parameterized by $\boldsymbol{\theta}$. Assume now that for some $\boldsymbol{\theta_0} \in \boldsymbol{\Theta}$, $M_t(\boldsymbol{\theta_0})$ holds for all t with $\boldsymbol{\theta_0}$ unknown. Usual methods, e.g. the ordinary Kalman filter, estimate the density $p(\mathbf{X_t}|\boldsymbol{\theta}, \mathbf{D}^t)$. Starting with an initial prior density $p(\boldsymbol{\theta}|\mathbf{D}^0)$ for $\boldsymbol{\theta}$ we can recursively calculate

$$p(\boldsymbol{\theta}|\mathbf{D}^t) \propto p(\boldsymbol{\theta}|\mathbf{D}^{t-1}) p(\mathbf{Y}_t|\boldsymbol{\theta}, \mathbf{D}^{t-1})$$

and use the *conditional* state density $p(\mathbf{X}_t|\boldsymbol{\theta}, \mathbf{D}^t)$ to obtain the *unconditional* state density

$$p(\mathbf{X}_t|\mathbf{D}^t) = \int_{\boldsymbol{\Theta}} p(\mathbf{X}_t|\boldsymbol{\theta}, \mathbf{D}^t) p(\boldsymbol{\theta}|\mathbf{D}^t) d\boldsymbol{\theta}$$

In practice the densities for $\boldsymbol{\theta}$ are mass functions since the parameter space is usually assumed to be a finite discrete set $\boldsymbol{\Theta} = \{\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_r}\}$. The resulting model is called a **multi-process class I** model by West & Harrison (1997). The final predictive density is obtained as a discrete probability mixture of the predictive densities associated to the components $M_t(\boldsymbol{\theta}_i)$, $i = 1, \ldots, r$, which in this case have been assumed to be Gaussian. **Multi-process class II** models are defined by West & Harrison (1997) as a generalization of **multi-process class I** models to situations in which the parameter vector $\boldsymbol{\theta}_t$ is allowed to be time varying.

5.2 Combination of forecasts

In order to reduce the dependence of the obtained predictions from an underlying forecasting model, a technically different solution is to calculate forecasts from different models and then obtain the final prediction by linearly combining the forecasts singularly obtained

$$f^{\star}(n,k) = \sum_{i=1}^{r} w_i f_i(n,k)$$

where the $f_i(n,k)$, $i = 1, \ldots, r$, are the forecasts to be combined and the w_i are the combining weights. More details on this topic will be given in the next section.

Empirical evidence and theoretical results both suggest that in many cases forecast accuracy could be sensibly improved through the combination of multiple individual forecasts. A review of forecasts combination methods has been given by Clemen (1989) and, more recently, by Diebold & Lopez (1996). Since combining methods do not involve a formal procedure for identifying the underlying data generating process, theoretical variance expressions are not easily derived unless simple combining functions are considered. However, simulation studies can be used to assess and compare the accuracy of different methods (e.g. Taylor & Bunn, 1999). The results obtained show that simple combination methods often work reasonably well relative to more complex combinations. In this section we will give a short account of the most common procedures for combining forecasts and of the problems associated to their practical implementation.

The first method we consider is based on the application of simple ordinary least squares regression techniques (OLS). In the followings we will refer to this method simply as to the *OLS combining method*. Let

$$\mathbf{f}_{t+k} = \{1, f_1(t,k), \ldots, f_r(t,k)\}$$

be a vector containing the r constituent forecasts to be combined and

$$\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_r\}'$$

a vector of weights. We can write

$$Y_{t+k} = \mathbf{f}_{t+k}\boldsymbol{\beta} + \epsilon_{t+k} \tag{34}$$

where ϵ_{t+k} is the error term, i.i.d with zero mean and uncorrelated with $\mathbf{f}_{t+k}$. Furthermore, let $\mathbf{F}^{t+k-1}$ be the $(t+k-1)\times(r+1)$ matrix of past forecasts. Ordinary least squares can be used to obtain an unbiased estimate of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = [(\mathbf{F^{t+k-1}})'\mathbf{F^{t+k-1}}]^{-1}(\mathbf{F^{t+k-1}})'\mathbf{Y}^{t+k-1}. \tag{35}$$

The predictor of Y_{t+k}, given $\mathbf{f}_{t+k}$, can then be expressed as

$$y(t,k) = \mathbf{f}_{t+k}\hat{\boldsymbol{\beta}} \tag{36}$$

From the unbiasedness of $\hat{\boldsymbol{\beta}}$ it follows that the PMSE of the predictor (36) is given by

$$E[Y_{t+k} - y(t,k)]^2 = \sigma_\epsilon^2\{\mathbf{f}_{t+k}[\mathbf{F^{t+k-1}}(\mathbf{F^{t+k-1}})']\mathbf{f}'_{t+k} + 1\}$$

where $\sigma_\epsilon^2 = var(\epsilon_t)$.

The *minimum variance* method (Bates and Granger, 1969) is equivalent to the OLS method with the restriction that $\beta_0 = 0$ and the coefficients sum to unity, $\sum_{i=1}^{r}\beta_i = 1$. An estimate of $\boldsymbol{\beta}$ can be obtained estimating the regression model in (34) subject to the constraint $\mathbf{R}\boldsymbol{\beta} = \boldsymbol{\gamma}$, where $\mathbf{R}$ is a $2\times(r+1)$ matrix such that $R_{11} = 1, R_{2j} = 1$ for $j > 1$ and with zeros elsewhere, $\boldsymbol{\gamma} = \{0 \quad 1\}'$. The constrained minimization problem is solved applying the technique of Lagrange Multipliers (Taylor and Bunn, 1999) to give the estimator

$$\hat{\boldsymbol{\beta}}^\star = \hat{\boldsymbol{\beta}} - [(\mathbf{F^{t+k-1}})'\mathbf{F^{t+k-1}}]^{-1}\mathbf{R}'\{\mathbf{R}[\mathbf{F^{t+k-1}}(\mathbf{F^{t+k-1}})']^{-1}\mathbf{R}'\}^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}).$$

Let $\mathbf{M} = [\mathbf{R}((\mathbf{F^{t+k-1}})'\mathbf{F^{t+k-1}})^{-1}\mathbf{R}']^{-1}$. The PMSE of the predictor

$$y^\star(t,k) = \mathbf{f}_{t+k}\hat{\boldsymbol{\beta}}^\star \tag{37}$$

is

$$\begin{aligned} E[Y_{t+k} - y^\star(t,k)]^2 &= \sigma_\epsilon^2\{\mathbf{f}_{t+k}[\mathbf{I} - ((\mathbf{F^{t+k-1}})'\mathbf{F^{t+k-1}})^{-1}\mathbf{R}'(\mathbf{I} - \sigma_\epsilon^{-2}\mathbf{M}\boldsymbol{\Gamma}\boldsymbol{\Gamma}') \\ &\quad \mathbf{M}\mathbf{R}]((\mathbf{F^{t+k-1}})'\mathbf{F^{t+k-1}})^{-1}\mathbf{f}'_{t+k} + 1\} \end{aligned}$$

with bias and variance of the prediction error given respectively by

$$\begin{aligned} var(Y_{t+k} - y^\star(t,k)) &= \sigma_\epsilon^2\{\mathbf{f}_{t+k}[\mathbf{I} - ((\mathbf{F^{t+k-1}})'\mathbf{F^{t+k-1}})^{-1}\mathbf{R}'\mathbf{M}\mathbf{R}] \\ &\quad ((\mathbf{F^{t+k-1}})'\mathbf{F^{t+k-1}})^{-1}\mathbf{f}'_{t+k} + 1\} \qquad (38) \\ E(Y_{t+k} - y^\star(t,k)) &= \mathbf{f}_{t+k}((\mathbf{F^{t+k-1}})'\mathbf{F^{t+k-1}})^{-1}\mathbf{R}'\mathbf{M}\boldsymbol{\Gamma} \qquad (39) \end{aligned}$$

where $\mathbf{\Gamma} = (\mathbf{R}\boldsymbol{\beta} - \boldsymbol{\gamma})$; Evaluation of the bias expression in eq. (39) is not an easy task and, at the same time, it is necessary for an assessment of the uncertainty associated to $\boldsymbol{\beta}^{\star}$. A common practice is to assume that the forecasts are unbiased i.e. that the restrictions on the unknown coefficients are verified and $\mathbf{R}\boldsymbol{\beta} = \boldsymbol{\gamma}$.

An alternative to the previous procedures is to calculate a simple average of the forecasts. This method can be also considered as a restricted least squares problem with the restriction matrix $\mathbf{R}$ equal to the identity matrix and the restriction vector $\boldsymbol{\gamma}$ having a first entry of zero and all the rest equal to $1/m$ (Aksu and Gunter, 1992). The variance of the prediction error is equal to the variance of the error term ϵ_{t+k}.

Analytical derivation of the PMSE for restricted least squares methods requires some sort of assumption on the bias of the predictor. Alternatively, it is possible to consider empirical procedures which make use of past within sample forecast errors at different lead times (e.g. Gardner, 1988; Taylor and Bunn, 1999). A limit of these procedures is that they do not produce accurate estimates of prediction intervals if only a limited number of forecast errors is available.

5.3 Other approaches

A third alternative to assuming the existence of a single true model is to use time-varying parameters models. These are based on the coexistence of several *local* models which hold at different time points. In this way, we are still relying on the assumption that the data have been generated by a given model structure but we make the model more flexible by allowing its parameters to vary through time, according to a given probabilistic law. Basically, in order to obtain an estimate of the time varying parameters, two main different approaches are feasible. The first relies on the application of Kalman filtering techniques after that an appropriate transition equation has been specified (e.g. Young, 1984). The second approach uses non-parametric smoothers, such as smoothing splines and local polynomial regression techniques, to estimate the variations in the parameters (e.g. Hastie & Tibshirani, 1993).

The strategy of *scenario forecasting* is widely diffused among practitioners in order to face model selection problems. It consists in calculating forecasts conditional on a number of different scenarios, which are reflected in different future values of the explanatory variables or in different model structures. This approach can be appropriate in cases in which the analyst is interested more in quantifying the effects of future extreme events rather than in forecasting the value that the process is likely to assume in the next time periods.

Finally, some suggest using different models for different lead times. It has been estabilished empirically (e.g. Gersch and Kitagawa, 1983) that the model

which works best for short term forecasting may not be so good for longer times.

6 CONCLUDING REMARKS

It should always be clear that all forecasts are based on more or less restrictive assumptions. These assumptions are usually motivated by the prior knowledge that the forecaster has of the system which has to be modeled and by the results of the exploratory analysis performed on the data (plots, detection of outliers, ACF and PACF, spectral analysis, linearity tests, etc...). Furthermore, even if the assumptions made are true, the forecasting model can only be considered as an useful approximate representation of the data generating process.

Provided that the identified model gives a good representation of the underlying phenomenon, the quality of forecasts will then depend on the *amount* of information given as an input to the model and on the *quality* of this information. In general, a multivariate forecasting method with many explanatory variables will not necessarily perform better than univariate methods that use, as only source of information, past realizations of the process which has to be forecast. When a model contains explanatory variables, forecasts can only be made conditional on future values of these variables which have themselves to be forecast. This may necessitate the construction of a model introducing an additional source of uncertainty in the forecasting process which, in many cases, can be difficult to control. Also, the inclusion of unnecessary explanatory variables, as well as the exclusion of relevant variables, can have a dramatic effect on the forecasting performance. The identification of all the explanatory variables which are relevant for reconstructing the future paths of the dependent variables is a delicate and, often, not easy task. On the other hand, artificially increasing the number of regressors could improve the fit but lead to poorer out of sample forecasts. So a badly misspecified multivariate model could lead to worse forecasts than a naive time series model. The above considerations suggest that a multivariate model can be even more vulnerable to misspecification than an univariate one.

Similar issues arise when choosing between a complicated non-linear model and a simple linear ARMA type structure. First, if there is no evidence of non-linear behaviour, a non-linear model will be obviously not appropriate. Second, if the functional form of a non-linear relation cannot be reliably identified, it could be advisable to resort to a simpler linear model that, in any case, can be used to obtain an approximate description of the phenomenon. Differently, a badly misspecified non-linear model will produce misleading results. Non-parametric methods can be a valuable tool for identifying the shape of the underlying non-linear relationship.

Finally, where this is possible, a good practice would be that of validating the forecasting model calculating *genuine* out-of-sample forecasts and not within sample prediction errors which tend to produce an overoptimistic measure of the forecast uncertainty.

REFERENCES

Aksu, C. & Gunter, S. I. An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts, *International Journal of Forecasting*, 1992, **8**, 27-43.

Amendola, A. & Storti, G. A threshold model for the rainfall-flow non-linearity, *Proceedings of the conference "SCO 99: Complex Models and Computationally Intensive Methods for Estimation and Forecasting"*, Venezia, 27-29 September 1999, 179-184.

Bates, J. M. & Granger C. W. J. The combination of forecasts, *Operational Research Quarterly*, 1969, **20**, 451-468.

Bollerslev, T., Generalized autoregressive conditional heteroscedasticity *Journal of Econometrics*, 1986, **31**, 307-327.

Bougerol, P. & Picard, N. Stationarity of GARCH processes and some non-negative time series, *Journal of Econometrics*, 1992, **52**, 115-127.

Box, G. E. P. & Jenkins, G. M. *Time series analysis, forecasting and control*, Holden Day, San Francisco, 1970.

Brockwell, P. J. & Davis, R. A. *Time series: Theory and Methods*, Springer, New York, 1987.

Carlin, B. P., Polson, N. G. & Stoffer, D. S. A Monte Carlo approach to nonnormal and nonlinear state-space modeling, *Journal of the American Statistical Association*, 1992, **87**, 493-500.

Catlin, D. E. *Estimation, Control, and the Discrete Kalman Filter*, Spinger-Verlag, Berlin, 1989.

Chatfield, C. Calculating Forecast Intervals, *Journal of Business and Economic Statistics*, 1993, **11**, 121-144.

Chatfield, C. Forecasting in the 1990s, *The Statistician*, 1997, **46**, 461-473.

Chen, R. & Hafner, C. Nonlinear time series analysis, in *XploRE, an Interactive Statistical Computing Environment*, ed. Härdle, W., Klinke, S. & Turlach, B., Springer-Verlag, Heidelberg, 1995.

Claps, P., Rossi, F. & Vitale, C. Conceptual-stochastic modeling of seasonal runoff usinfg autoregressive moving average models and different scales of aggregation, *Water Resources Research*, 1993, **29**, 2545-2559.

Clemen, R. T. Combining forecasts:a review and annotated bibliography, *International Journal of Forecasting*, 1989, **5**, 559-583.

Davison, A C. & Hinkley, V. *Bootstrap Methods and Their Application*, Cambridge University Press, 1997.

Diebold, F. X. & Lopez J. A. Forecast Evaluation and Combination, Technical Working Paper 192, National Bureau of Economic Research, Cambridge MA, 1996

Diebolt, J. & Guegan, D. Probabilistic properties of the general non-linear Markovian process of order one and applications to time series modeling, Technical Report, Laboratoire de Statistique, Universitè Paris XIII, 1991.

Draper, D. Assessment and propagation of model uncertainty (with discussion), *Journal of the Royal Statistical Society B*, 1995, **57**, 45-97.

Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap*, Chapman & Hall, London, 1993.

Engle, R. F., Granger, C. W. J.,Rice, J. & Weiss, A. Semiparametric estimates of the relation between weather and electricity sales, *Journal of the American Statistical Association*, 1986, **81**, 310-320.

Fan, J. & Yao, Q. Efficient estimation of conditional variance functions in stochastic regression, *Biometrika*, 1998, **85**, 645-660.

Gallant, A. R. & Tauchen, G. E. Semiparametric estimation of conditional constrained heterogeneous processes: asset pricing applications, *Econometrica*, 1989, **57**, 1091-1120.

Gardner, Jr., E. S. A simple method for computing prediction intervals for time series forecasts, *Management Science*, 1988, **34**, 541-546.

Gersch, W. & Kitagawa, G. The prediction of time series with trends and seasonalities, *Journal of Busines and Economic Statistics*, 1983, **1**, 253-264.

Gilks, W. R., Richardson, S. & Spiegehalter, D. J. *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London, 1996.

Giordano, F. & Perna, C. The hidden layer's size in feed-forward neural networks: a statistical point of view, *Proceedings of the conference "SCO 99: Complex Models and Computationally Intensive Methods for Estimation and Forecasting"*, Venezia, 27-29 September 1999, 95-100.

Granger, C. W. J. & Andersen, A. P. *An Introduction to Bilinear Time Series Models*, Vandenhoek and Ruprecht, Gottingen, 1978.

Hamilton, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 1989, **57**, 357-384.

Härdle, W. *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 1990.

Härdle,W., Lütkepohl, H. & Chen, R. A review of non-parametric time series analysis, *International Statistical Review*, 1997, **65**, 49-72.

Härdle, W. & Tsybakov, A. B. Local polynomial estimators of the volatility function, *Journal of Econometrics*, 1997, **81**, 223-242.

Härdle,W. & Vieu, P. Kernel regression smoothing of time series, *Journal of Time Series Analysis*, 1992, **13**, 209-232.

Harvey, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 1989.

Hastie, T. and Tibshirani, R. Varying-coefficient models, *Journal of the Royal Statistical Society B*, 1993, **55**, 757-796.

Kitagawa, G. Non-Gaussian state space modeling of nonstationary time series, *Journal of the American Statistical Association*, 1987, **82**, 1032-1063.

Kalman, R. E. A new approach to linear filtering and prediction problems, *Journal of Basic Engineering (ASME)*, 1960, **82D**, 35-45.

Künsch, H. R. The jackknife and the bootstrap for general stationary observations, *Annals of Statistics*, 1989, **17**, 1217-1241.

La Rocca, M. & Vitale, C. Boostrap inference in bilinear models, *Proceedings of the conference "SCO 99: Complex Models and Computationally Intensive Methods for Estimation and Forecasting"*, Venezia, 27-29 September 1999, 64-79.

Li, W. K. & Lam, K. Modeling asymmetry in stock returns by a threshold ARCH model, *The Statistician*, 1995, **44**, 333-341.

Lütkepohl, H. *Introduction to Multiple Time Series Analysis* (2^{nd} ed.), Springer-Verlag, Heidelberg, 1993.

Masry, E. & Tjøstheim, D. Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality, *Econometric Theory*, 1995, **11**, 258-289.

Nisio, M. On polynomial approximation for strictly stationary processes, *J. Math Soc. Jpn*, 1960, **12**, 207-226.

Phillips, P. C. B. The sampling distribution of forecasts from a first order autoregression, *Journal of Econometrics*, 1979, **9**, 241-261.

Piccolo, D. & Vitale, C. *Metodi statistici per l'analisi economica*, Il Mulino, Bologna, 1981.

Politis, D. N. & Romano, J. P. The stationary Bootstrap, *Journal of the American Statistical Association*, 1994, **89**, 1303-1313.

Priestley, M. B. State-dependent models: a general approach to non-linear time series analysis, *Journal of Time Series Analysis*, 1980, **1**, 47-71.

Reinsel, G. *Elements of Multivariate Time Series Analysis* (2^{nd} ed.), Springer, New York, 1997.

Robinson, P.M. Non-parametric estimation for time series models, *Journal of Time Series Analysis*, 1983, **4**, 185-208.

Rosenblatt, M. *Markov processes, structure and asymptotic behaviour*, Journal of the American Statistical Association, Springer, New York, 1971.

Rudolph, A. A central limit theorem for random coefficient autoregressive models and ARCH/GARCH models, *Advances in Applied Probability*, 1998, **30**, 113-121.

Shao, J. & Tu, D. *The Jackknife and the Bootstrap*, Springer, New York, 1995.

Shumway, R. H., Azari, A. S. & Pawitan, Y. Modeling mortality fluctuations in Los Angeles as functions of pollution and weather effects, *Environmental Research*, 1988, **45**, 224-241.

Storti, G., Furcolo, P. & Villani, P. Uncertainty asessment of extreme rainfall forecast, *Paper presented at the XXIII EGS General Assembly*, Nice, 1998.

Tanizaki, H. & Mariano, R. S. Non-linear and non-Gaussian state-space modeling with Monte Carlo simulations, *Journal of Econometrics*, 1998, **83**, 263-290.

Taylor, J. W. & Bunn, D. W. Investigating improvements in the accuracy of prediction intervals for combinations of forecasts: a simulation study, *International Journal of Forecasting*, 1999, **15**, 325-339.

Thombs, L. A. & Schucany, W. R. Bootstrap prediction intervals for autoregression, *Journal of the American Statistical Association*, 1990, **85**, 486-492.

Tjøstheim, D. Non-linear time series: a selective review, *Scandinavian Journal of Statistics*, 1994, **21**, 97-130.

Tong, H. *Non-linear time series. A dynamical system approach*, Oxford University Press, London, 1990.

Tong, H. & Lim K. S. Threshold autoregression, limit cycles and cyclical data (with discussion), *Journal of the Royal Statistical Society B*, 1980, **42**, 245-292.

Tsay, R. Testing and Modelling Multivariate Threshold Models, *Journal of the American Statistical Association*, 1998, **93**, 1188-1202.

West, M. & Harrison, J. *Bayesian Forecasting and Dynamic Models (2nd edition)*, Springer, New York, 1997.

Young, P. C. *Recursive Estimation and Time-Series Analysis (An Introduction)*, Springer-Verlag, Berlin, 1984.