UNIVERSITY OF GÖTEBORG

Department of Statistics

RESEARCH REPORT 1992:1

ISSN 0349-8034

ASPECTS OF MODELLING NONLINEAR TIME SERIES

by

Timo Teräsvirta, Dag Tjøstheim and Clive W J Granger

Statistiska institutionen
Göteborgs Universitet
Viktoriagatan 13
S 411 25 Göteborg
Sweden

ASPECTS OF MODELLING NONLINEAR TIME SERIES

by

Timo Teräsvirta*, Dag Tjøstheim** and Clive W.J. Granger***

* Research Institute of the Finnish Economy,
Lönnrotinkatu 4 B, SF-00120 Helsinki, Finland

** Department of Mathematics, University of Bergen,
N-5000 Bergen, Norway

*** Department of Economics, University of California,
San Diego, La Jolla, CA 92093-0508, USA

First draft: November 1991
This version: January 1992

Acknowledgements. The work for this paper was originated when TT and DT were visiting University of California, San Diego. They wish to thank the economics and mathematics departments, respectively, of UCSD for their hospitality and John Rice and Murray Rosenblatt in the latter in particular. The research of TT was also supported by University of Göteborg and a grant from the Yrjö Jahnsson Foundation. DT acknowledges financial support from the Norwegian Council for Research and CWJG from NSF, Grant SES 9023037.

1. INTRODUCTION

It is common practice for economic theories to postulate non-linear relationships between economic variables, production functions being an example. If a theory suggests a specific functional form, econometricians can propose estimation techniques for the parameters, and asymptotic results, about normality and consistency, under given conditions are known for these estimates, see e.g. Judge et. al. (1985) and White (1984) and Gallant (1987, chapter 7). However, in many cases the theory does not provide a single specification or specifications are incomplete and may not capture the major features of the actual data, such as trends, seasonality or the dynamics. When this occurs, econometricians can try to propose more general specifications and tests of them. There are clearly an immense number of possible parametric nonlinear models and there are also many nonparametric techniques for approximating them. Given the limited amount of data that is usually available in economics it would not be appropriate to consider many alternative models or to use many techniques. Because of the wide possibilities the methods and models available to analyze non-linearities are usually very flexible so that they can provide good approximations to many different generating mechanisms. A consequence is that with fairly small samples the methods are inclined to over-fit, so that if the true mechanism is linear, say, with residual variance σ^2, the fitted model may appear to find nonlinearity and the estimated residual variance is less than σ^2. The estimated model will then be inclined to forecast badly in the post-sample period. It is therefore necessary to have a specific research strategy for modelling non-linear relationships between time series. In this chapter the modelling process concentrates on a particular situation, where there is a single dependent variable y_t to be explained and $\underline{x}_t$ is a vector of exogenous variables. Let I_t be the information set

$$I_t = \{ y_{t-j}, j > 0; \underline{x}_{t-i}, i \geq 0 \} \tag{1.1}$$

and denote all of the variables (and lags) used in I_t by $\underline{w}_t$. The modelling process will then attempt to find a satisfactory approximation for $f(\underline{w}_t)$ such that

$$E[y_t | I_t] = f(\underline{w}_t). \tag{1.2}$$

If the error is

$$\varepsilon_t = y_t - f(\underline{w}_t)$$

then in some cases a more parsimonious representation will specifically include lagged ε's in f( ).

The strategy proposed is:

(i) Test y_t for linearity, using the information I_t. As there are many possible forms of nonlinearity it is likely that no one test will be powerful against them all, so several tests may be needed.

(ii) If linearity is rejected, consider a small number of alternative parametric models and/or nonparametric estimates. Linearity tests may give guidance as to which kind of nonlinear models to consider.

(iii) These models should be estimated in-sample and compared out-of-sample. The properties of the estimated models should be checked. If a single model is required, the one that is best out-of-sample may be selected and re-estimated over all available data.

The strategy is by no means guaranteed to be successful. For example, if the nonlinearity is associated with a particular feature of the data, but if this feature does not occur in the post-sample evaluation period, then the nonlinear model may not perform any better than a linear model.

Section 2 of the chapter briefly considers some parametric models, Section 3 discusses tests of linearity, Section 4 reviews specification of nonlinear models, Section 5 considers estimation and Section 6 evaluation of estimated models. Section 7 contains an example and section 8 concludes. This survey largely deals with linearity in the conditional mean, which occurs if $f(\underline{w}_t)$ in (1.1) can be well approximated by some linear combination $\underline{\varphi}' \underline{w}_t$ of the components of $\underline{w}_t$. It will generally be assumed that $\underline{w}_t$ contains lagged values of y_t plus, possibly, present and lagged values of $\underline{x}_t$ including 1. This definition avoids the difficulty of deciding whether or not processes having forms of heteroskedasticity that involve explanatory or lagged variablels, such as ARCH, are non-linear. It is clear that some tests of linearity will be confused by these types of heteroskedasticity. Recent surveys of some of the topics considered here include Tong (1990) for univariate time series, Härdle (1990) for non-parametric techniques, Brock and Potter (1992) for linearity testing and Granger and Teräsvirta (1992).

There has recently been a lot of interest, particularly by economic theorists in chaotic processes, which are deterministic series which have some of the linear properties of familiar

stochastic processes. A well known example is the "tent-map" $y_t = 4y_{t-1}(1-y_{t-1})$, which, with a suitable starting value in (0,1), generates a series with all autocorrelations equal to zero and thus a flat spectrum, and so may be called a "white chaos", as a stochastic white noise also has these properties. Economic theories can be constructed which produce such processes as discussed in Chen and Day (1992). Econometricians are unlikely to expect such models to be relevant in economics, having a strong affiliation with stochastic models and so far there is no evidence of actual economic data having been generated by a deterministic mechanism. A difficulty is that there is no statistical test which has chaos as a null hypothesis, so that non-rejection of the null could be claimed to be evidence in favour of chaos. For a discussion and illustrations, see Liu et. al. (1991). However, a useful linearity test has been proposed by Brock et. al. (1987), based on chaos theory, whose properties are discussed in section 3.2.

The hope in using nonlinear models is that better explanations can be provided of economic events and consequently better forecasts. If the economy were found to be chaos, and if the generating mechanism can be discovered, using some learning model say, then forecasts would be effectively exact, without any error.

2. TYPES OF NONLINEAR MODELS

2.1. Models from economic theory

Theory can both suggest possibly sensible nonlinear models or can consider some optimizing behaviour, with arbitrary assumed cost or utility functions, to produce a model. An example is a relationship of the form

$$y_t = \min(\underline{\varphi}' \underline{w}_t, \underline{\theta}' \underline{w}_t) + \varepsilon_t \qquad (2.1)$$

so that y_t is the smallest of a pair of alternative linear combinations of the vector of variables used to model y_t. This model arises from a disequilibrium analysis of some simple markets, with the linear combinations representing supply and demand curves, for more discussion see Quandt (1982) and Maddala (1986).

If we replace the "min condition" by another variable z_{t-d} which may also be one of the elements of w_t but not 1, we may have

$$y_t = \varphi' w_t + \theta' w_t F(z_{t-d}) + \varepsilon_t \qquad (2.2)$$

where $F(z_{t-d}) = 0, z_{t-d} \leq c, F(z_{t-d}) = 1, z_{t-d} > c$. This is a switching regression model with switching variable z_{t-d} where d is the delay parameter; see Quandt (1983). In univariate time series analysis (2.2) is called a two-regime threshold autoregressive model; see e.g. Tong (1990). Model (2.2) may be generalized by assuming a continuum of regimes instead of only two. This can be done for instance by defining

$$F(z_{t-d}) = (1 + \exp\{-\gamma(z_{t-d} - c)\})^{-1}, \gamma > 0$$

in (2.2). Maddala (1977, p. 396) already proposed such a generalization which is here called a logistic smooth transition regression model. F may also have a form of a probability density rather than cumulative distribution function. In the univariate case this would correspond to the exponential smooth transition autoregressive model (Teräsvirta, 1990a) or its well-known special case, the exponential autoregressive model (Haggan and Ozaki, 1981). The transition variable may represent changing political or policy regimes, high inflation versus low, upswings of the business cycle versus the downswings and so forth. These switching models or their smooth transition counterparts occur frequently in theory which, for example, suggests changes in relationships when there is idle production capacity versus otherwise or when unemployment is low versus high. Aggregation considerations suggest that a smooth transition regression model may often be more sensible than the abrupt change in (2.2).

Some theories lead to models that have also been suggested by time series statisticians. An example is the bivariate non-linear autoregressive model described as a "prey-predator" model by Desai (1984) taking the form

$$\Delta y_{1t} = -a + b \exp(y_{2t})$$

$$\Delta y_{2t} = c + b \exp(y_{1t})$$

where y_1 is the logarithm of the share of wages in national income and y_2 is the logarithm of the employment rate. Other examples can be found in the conference volume Chen and Day (1992). The fact that some models do arise from theory justifies their consideration but it does not imply that they are necessarily superior to other models that currently do not arise from economic theory.

2.2. Models from time series theory

The linear autoregressive, moving average and transfer function models have been popular in the time series literature following the work by Box and Jenkins (1970) and there are a variety of natural generalizations to non-linear forms. If the information set being considered is

$$I_t = \{ y_{t-j}, j = 1,...,q, \underline{x}_{t-i}, i = 0,...,q \}$$

denote ε_t the residual from y_t explained by I_t and let e_{kt} be the residual from x_{kt} explained by I_t (excluding x_{kt} itself). The components of the models considered in this section are non-linear functions of components such as $g(y_{t-j})$, $h(x_{k,t-i})$, $G(\varepsilon_{t-j})$, $H(e_{k,t-i})$ plus cross-products such as $y_{t-j}\, x_{k,t-i}$, $y_{t-j}\, \varepsilon_{t-i}$, $x_{a,t-j}\ e_{b,t-i}$ *or* $\varepsilon_{t-j}\, e_{k,t-i}$. A model would string together several such components, each with a parameter. For a given specification, the model is linear in the parameters so they can be easily estimated by OLS. The big questions are about specification of the model, what components to use, what functions and what lags. There are so many possible components and combinations that the "curse of dimensionality" soon becomes apparent, so that choices of specification have to be made. Several classes of models have been considered. They include

(i) nonlinear autoregressive, involving only functions of the dependent variable. Typically only simple mathematical functions have been considered (such as cosine, sign, modulus, integer powers, logarithm of modulus or ratios of low order polynomials);

(ii) nonlinear transfer functions, using functions of the lagged dependent variable and current and lagged explanatory variables, usually separately;

(iii) bilinear models, $y_t = \sum_{j,k} \beta_{jk}\, y_{t-j}\, \varepsilon_{t-k}$ + similar terms involving products of a component of $\underline{x}_t$ and a lagged residual of some kind. This can be thought of as one equation of a multivariate bilinear system as considered by Stensholt and Tjøstheim(1987);

(iv) nonlinear moving averages, being sums of functions of lagged residuals ε_t, $\underline{e}_t$;

(v) doubly stochastic models contain the cross-products between lagged y_t and current and lagged components of x_{kt} or a random parameter process and are

discussed in Tjøstheim (1986).

Most of the models are augmented by a linear autoregressive term. There has been little consideration of mixtures of these models. Because of difficulty of analysis lags are often taken to be small. Specifying the lag structure in nonlinear models is discussed in section 4.

A number of results are available for some of these models, such as stability for simple nonlinear autoregressive models (Lasota & Mackey, 1987), stationarity and invertibility of bilinear models or the autocorrelation properties of certain bilinear systems but are often too complicated to be used in practice. To study stability or invertibility of a specific model it is recommended that a long simulation be formed and the properties of the resulting series be studied. There is not a lot of experience with the models in a multivariate setting and little success in their use has been reported. At present they cannot be recommended for use compared to the smooth transition regression model of the previous section or the more structured models of the next section. A simple nonlinear autoregressive or bilinear model with just a few terms may be worth considering from this group.

2.3. Flexible statistical parametric models

A number of important modelling procedures concentrate on models of the form

$$y_t = \underline{\beta}' \underline{w}_t + \sum_{j=1}^{p} \alpha_j \varphi_j (\underline{\gamma}_j' \underline{w}_t) + \varepsilon_t \qquad (2.4)$$

where $\underline{w}_t$ is a vector of past y_t values and past and present values of a vector of explanatory variables $\underline{x}_t$ plus a constant. The first component of the model is linear and the $\varphi_j(x)$ are a set of specific functions in x, examples being:

(i) power series, $\varphi_j(x) \equiv x^j$ (x is generally not a lag of y) ;

(ii) trigonometric, $\varphi(x)$ = sin x or cos x, (2.4) augmented by a quadratic term $z_t' A z_t$ gives the flexible function forms discussed by Gallant (1981);

(iii) $\varphi_j(x) = \varphi(x)$ for all j, where $\varphi(x)$ is a "squashing function" such as a

probability density function or the logistic function $\varphi(x) = (1+exp(-x))^{-1}$. This is a neural network model, which has been used successfully in various fields, especially as a learning model, see e.g. White (1989);

(iv) if $\varphi_j(x)$ is estimated non-parametrically, by a "super-smoother", say, the method is that of "projection-pursuit", as briefly described in the next section.

The first three models are dense, in the sense that theorems exist suggesting that any well-behaved function can be approximated arbitrarily well by a high enough choice of p, the number of terms in the sum, for example Stinchcombe and White (1989). In practice, the small sample sizes available in economics limit p to a small number, say one or two, to keep the number of parameters to be estimated at a reasonable level. In theory p should be chosen using some stopping criterion or goodness-of-fit measure. In practice a small, arbitrary value is usually chosen, or some simple experimentation undertaken. These models are sufficiently structured to provide interesting and probably useful classes of nonlinear relationships in practice. They are natural alternatives to nonparametric and semiparametric models. A nonparametric model, as discussed in section 2.5 produces an estimate of a function at every point in the space of explanatory variables by using some smoother, but not a specific parametric function. The distinction between parametric and nonparametric estimators is not sharp, as methods using splines or neural nets with an undetermined cut off value indicate. This is the case in particular for the restricted nonparametric models in section 6.

2.4. State-space, time-varying parameter and long-memory models

Priestley (1988) has discussed a very general class of models for a system taking the form:

$$\underline{Y}_t = \sum \varphi_j(\underline{x}_{t-1})\,\underline{Y}_{t-j} + \mu(\underline{x}_{t-1}) + \varepsilon_t$$

(moving average terms can also be included) where $\underline{Y}_t$ is a k×1 stochastic vector and $\underline{x}_t$ is a "state-variable" consisting of $\underline{x}_t = (\underline{Y}_t, \underline{Y}_{t-1}, \ldots, \underline{Y}_{t-k+1})$ and which is updated by a Markov system,

$$\underline{x}_{t+1} = \mu(\underline{x}_t) + \underline{F}(\underline{x}_t)\,\underline{x}_t + \varepsilon_{t+1}.$$

Here the φ 's and the components of the matrix $\underline{F}$ are general functions, but in practice will be approximated by linear or low-order polynomials. Many of the models discussed in section 2.2 can be embedded in this form. It is clearly related to the extended Kalman filter (see Anderson and Moore, 1979) and to time-varying parametric ARMA models, where the parameters evolve according to some simple AR model, see Granger and Newbold (1986, chapter 10). For practical use various approximations can be applied, but so far there is little actual use of these models with multivariate economic series.

For most of the models considered in section 2.2, the series are assumed to be stationary, but this is not always a reasonable assumption in economics. In a linear context many actual series are I(1), in that they need to be differenced in order to become stationary, and some pairs of variables are cointegrated, in that they are both I(1) but there exists a linear combination that is stationary. A start to generalizing these concepts to nonlinear cases has been made by Granger and Hallman (1991a,b). I(1) is replaced by a long-memory concept, cointegration by a possibly nonlinear attractor, so that y_t , x_t are each long-memory but there is a function $g(x)$ such that $y_t - g(x_t)$ is stationary. A nonparametric estimator for $g(x)$ is proposed and an example provided.

2.5. Nonparametric models

Nonparametric modelling of time series does not require an explicit model but for reference purposes it is assumed that there is the following model

$$y_t = f(\underline{y}_{t-1}, \underline{x}_{t-1}) + g(\underline{y}_{t-1}, \underline{x}_{t-1})\,\varepsilon_t \tag{2.5}$$

where $\{y_t, x_t\}$ are observed with $\{x_t\}$ being exogeneous, and where $\underline{y}_{t-1} = (y_{t-i_1}, \ldots, y_{t-i_p})$ and $\underline{x}_{t-1} = (x_{t-j_1}, \ldots, x_{t-j_p})$ are vectors of lagged variables, and $\{\varepsilon_t\}$ is a sequence of martingale differences with respect to the information set $I_t = \{y_{t-i}, i > 0; x_{t-i}, i > 0\}$. The joint process $\{y_t, x_t\}$ is assumed to be stationary and strongly mixing (cf. Robinson, 1983). The model formulation can be generalized to several variables and instantaneous transformation of exogeneous variables. There has recently for instance been a surge of interest in nonparametric modelling, for references see for instance Ullah (1989), Barnett et al. (1991) and Härdle (1990). The motivation is to approach the data with as much flexibility as possible not being restricted by the straitjacket of a particular class of parametric models. However, more observations are needed to obtain estimates of comparable variability. In

econometric applications the two primary quantities of interest are the conditional mean

$$M(\underline{y};\underline{x}) = M(y_1,\ldots,y_p;x_1,\ldots,x_q)$$

$$= E(y_t \mid y_{t-i_1} = y_1,\ldots,y_{t-i_p} = y_p; x_{t-j_1} = x_1,\ldots,x_{t-j_q} = x_q) \tag{2.6}$$

and the conditional variance

$$V(\underline{y};\underline{x}) = V(y_1,\ldots,y_p;x_1,\ldots,x_q)$$

$$= var(y_t \mid y_{t-i_1} = y_1,\ldots,y_{t-i_p} = y_p; x_{t-j_1} = x_1,\ldots,x_{t-j_q} = x_q)\,. \tag{2.7}$$

The conditional mean gives the optimal least squares predictor of y_t given lagged values $y_{t-i_1},\ldots,y_{t-i_p}; x_{t-j_1},\ldots,x_{t-j_q}$. Derivatives of $M(\underline{x};\underline{y})$ can also have economic interpretations (Ullah, 1989) and can be estimated nonparametrically. The conditional variance can be used to study volatility. For (2.5), $M(\underline{y},\underline{x}) = f(\underline{y},\underline{x})$ and $V(\underline{y},\underline{x}) = \sigma^2 g^2(\underline{y},\underline{x})$, where $\sigma^2 = E(\varepsilon_t^2)$. As pointed out in the Introduction, this survey mainly concentrates on $M(\underline{y};\underline{x})$ while it is assumed that $g(\underline{y};\underline{x}) \equiv 1$.

A problem of nonparametric modelling in several dimensions is the curse of dimensionality. As the number of lags and regressors increases, the number of observations in a unit volume element of regressor space can become very small, and it is difficult to obtain meaningful nonparametric estimates of (2.6) and (2.7). Special methods have been designed to overcome this obstacle, and they will be considered in sections 4 and 5.3. Applying these methods often results in a model which is an end product in that no further parametric modelling is necessary.

Another remedy to difficulties due to the dimension is to apply semiparametric models. These models usually assume linear and parametric dependence in some variables, and nonparametric functional dependence in the rest. The estimation of such models as well as restricted nonparametric ones will be considered in section 5.3.

3. TESTING LINEARITY

When parametric nonlinear models are used for modelling economic relationships, model specification is a crucial issue. Economic theory is often too vague to allow complete specification of even a linear, let alone a nonlinear model. Usually at least the specification

of the lag structure has to be carried out using the available data. As discussed in the Introduction, the type of nonlinearity best suited for describing the data may not be clear at the outset either. The first step of a specification strategy for any type of nonlinear model should therefore consist of testing linearity. As mentioned above it may not be difficult at all to fit a nonlinear model to data from a linear process, interpret the results and draw possibly erroneous conclusions. If the time series are short that may sometimes be successfully done even in situations in which the nonlinear model is not identified under the linearity hypothesis. There is more statistical theory available for linear than nonlinear models and the parameter estimation in the former models is generally simpler than in the latter. Finally, multi-step forecasting with nonlinear models is more complicated than with linear ones. Therefore the need for a nonlinear model should be considered before any attempt at nonlinear modelling.

3.1. Tests against a specific alternative

Since estimation of nonlinear models is generally more difficult than that of linear models it is natural to look for linearity tests which do not require estimation of any nonlinear alternative. In cases where the model is not identified under the null hypothesis of linearity, tests based on the estimation of the nonlinear alternative would normally not even be available. The score or Lagrange multiplier principle thus appears useful for the construction of linearity tests. In fact, many well-known tests in the literature are Lagrange multiplier (LM) or LM type tests. Moreover, some well-known tests like the test of Tsay (1986) which have been introduced as general linearity tests without a specific nonlinear alternative in mind can be interpreted as LM tests against a particular nonlinear model. This may not be surprising because those tests do not require estimation of a nonlinear model. Other tests, not built upon the LM principle, do exist and we shall mention some of them. Recent accounts of linearity testing in nonlinear time series analysis include Brock and Potter (1992), De Gooijer and Kumar (1991), Granger and Teräsvirta (1992, chapter 6) and Tong (1990, chapter 5). For small-sample comparisons of some of the tests, see Chan and Tong (1986), Lee et al. (1992), Luukkonen et al. (1988a) and Petruccelli (1990).

Consider the following nonlinear model

$$y_t = \underline{\varphi}' \underline{w}_t + f(\underline{\theta}, \underline{w}_t, \underline{v}_t) + u_t \tag{3.1}$$

where $\underline{w}_t = (1, y_{t-1}, ..., y_{t-p}, x_{t1}, ..., x_{tk})'$, $\underline{v}_t = (u_{t-1}, ..., u_{t-q})'$, $u_t = g(\underline{\psi}, \underline{\theta}, \underline{w}_t, \underline{v}_t)\,\varepsilon_t$, ε_t is a martingale difference process: $E(\varepsilon_t \mid I_t) = 0$, $cov(\varepsilon_t \mid I_t) = \sigma_\varepsilon^2$, where I_t is as in (1.1). It

follows that $E(u_t \mid I_t) = 0$ and $cov(u_t \mid I_t) = \sigma_\varepsilon^2 g^2(\psi, \theta, w_t, v_t)$. Assume that f and g are at least twice continuously differentiable with respect to the parameters $\underline{\theta} = (\theta_1, ..., \theta_m)'$ and $\underline{\Psi} = (\psi_1, ..., \psi_l)'$. Let $f(0, \underline{w}_t, \underline{v}_t) = 0$, so that the linearity hypothesis becomes $H_0 : \underline{\theta} = 0$.

To test this hypothesis assuming $g \equiv 1$ write the conditional (pseudo) logarithmic likelihood function as

$$
\begin{aligned}
& l(\underline{\theta}, \underline{\varphi}; y_T, w_T, w_{T-1}, ..., w_1, v_T, v_{T-1}, ..., v_1 | W_0, U_0) \\
& = \sum_1^T l_t(\underline{\theta}, \underline{\varphi}; y_t \mid w_t, ..., w_1, v_t, ..., v_1, W_0, U_0) \\
& = c - (T/2) \log \sigma_\varepsilon^2 - (1/2\,\sigma_\varepsilon^2) \sum_1^T u_t^2 .
\end{aligned}
$$

The relevant block of the score vector scaled by $1/\sqrt{T}$ becomes

$$
\frac{1}{\sqrt{T}} \frac{\partial l}{\partial \underline{\theta}} = (1/\sqrt{T}\,\sigma_\varepsilon^2) \sum u_t \underline{h}_t , \quad \underline{h}_t = \left(\frac{\partial u_t}{\partial \theta_1}, ..., \frac{\partial u_t}{\partial \theta_m} \right)' .
$$

This is the block that is nonzero under the null hypothesis. The information matrix is block diagonal so that the diagonal element conforming to σ^2 builds a separate block. Thus the inverse of the block related to θ and evaluated at H_0 becomes

$$
I_\theta = (T \sigma_\varepsilon^2)^{-1} \sum \tilde{\underline{h}}_t \tilde{\underline{h}}_t' - \sum \tilde{\underline{h}}_t \underline{w}_t' \left(\sum \underline{w}_t \underline{w}_t' \right)^{-1} \sum \underline{w}_t \tilde{\underline{h}}_t'
$$

where $\tilde{\underline{h}}_t$ is $\underline{h}_t$ evaluated at H_0; see e.g. Granger and Teräsvirta (1992, chapter 6). Setting $\tilde{\underline{u}} = (\tilde{u}_1, ..., \tilde{u}_T)'$ the test statistic, in obvious notation, has the form

$$
LM = \tilde{\underline{u}}' H (H' M_W H)^{-1} H' \tilde{\underline{u}} \tag{3.2}
$$

where $M_W = I - W(W'W)^{-1}W'$ and the vector $\tilde{u}$ consists of residuals from (3.1) estimated under H_0 and $g \equiv 1$. Under a set of assumptions which are moment conditions for (3.2), see White (1984, Theorem 4.25), (3.2) has an asymptotic $\chi^2(m)$ distribution. A practical way of carrying out the test is by ordinary least squares as follows:

(i) Regress y_t on $\underline{w}_t$, compute the residuals $\tilde{u}_t$ and the sum of squared residuals SSR_0.

(ii) Regress $\tilde{u}_t$ on $\underline{w}_t$ and $\underline{\tilde{h}}_t$,, compute the sum of squared desiduals SSR_1.

(iii) Compute

$$F(m, T-n-m) = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T-n-m)}$$

with $n=k+p+1$, which has an approximate F distribution under $\theta = 0$.

The use of an F test instead of the χ^2 test given by the asymptotic theory is recommended in small samples because of its good size and power properties, see Harvey (1990, p. 174-175).

As an example, assume $w_t = (1, \underline{\bar{w}}_t')'$ with $\underline{\bar{w}}_t = (y_{t-1}, ..., y_{t-q})'$ and $f = \underline{v}_t' \Theta \underline{\bar{w}}_t$ $= (\underline{v}_t \otimes \underline{\bar{w}}_t)'$ *vec* (Θ) so that (3.1) is a univariate bilinear model. Then $h_t = (\underline{v}_t \otimes \underline{\bar{w}}_t)$, $\underline{\tilde{h}}_t = (\underline{\tilde{v}}_t \otimes \underline{\bar{w}}_t)$ and (3.2) is a linearity test against bilinearity discussed in Weiss (1986) and Saikkonen and Luukkonen (1988).

In a few cases f in (3.1) factors as follows:

$$f(\underline{\theta}, \underline{w}_t) = (\underline{\theta}_1' \underline{w}_t) f_1(\theta_2, \underline{\theta}_3, \underline{w}_t) \qquad (3.3)$$

and $f_1(0, \underline{\theta}_3, \underline{w}_t) = 0$. Assume that θ_2 is a scalar whereas $\underline{\theta}_3$ may be a vector. This is the case for many nonlinear models such as smooth transition regression models. Vector v_t is dropped for simplicity. The linearity hypothesis can be expressed as H_{02}: $\theta_2 = 0$. However, H_{01}: $\theta_1 = 0$ is also a valid linearity hypothesis. This is an indication of the fact that (3.1) with (3.3) is only identified under the alternative $\theta_2 \neq 0$ but not under $\theta_2 = 0$. If we choose H_{02} as our starting-point, we may use the Taylor expansion

$$f_1(\theta_2, \underline{\theta}_3, w_t) = f_1(0, \underline{\theta}_3, \underline{w}_t) + b(0, \underline{\theta}_3, \underline{w}_t)'\theta_2 + R_2(\theta_2, \underline{\theta}_3, \underline{w}_t) \approx b_t\theta_2 . \qquad (3.4)$$

Assume furthermore that b_t has the form

$$b_t = \theta_{30} + \underline{\theta}_{31}' k(\underline{\bar{w}}_t) \qquad (3.5)$$

where $\underline{\theta}_{31}$ and $k(\underline{\bar{w}}_t)$ are r×1 vectors. Next replace f_1 in (3.3) by the first-order Taylor approximation at $\theta_2 = 0$

$$t(0, \underline{\theta}_3, w_t) = \{ \theta_{30} + \underline{\theta}_{31}' k(\underline{\bar{w}}_t) \} \theta_2 .$$

Then (3.3) becomes

$$\bar{f}\ (\theta, \underline{w}_t) = (\underline{\theta}'_1 \underline{w}_t)\ (\theta_{30} + \underline{\theta}'_{31} k(\bar{w}_t))\ \theta_2$$

$$= \bar{\underline{\psi}}'_1\ \underline{w}_t + k(\bar{w}_t)'\Psi \underline{w}_t$$

where $\bar{\underline{\psi}}_1 = \theta_2 \underline{\theta}_1 \theta_{30}$, $\Psi = \theta_2 \underline{\theta}_1 \underline{\theta}'_{31}$ and (3.1) has the form

$$y_t = (\underline{\varphi}_1 + \theta_2 \theta_{30} \underline{\theta}_1\ )' \underline{w}_t + k(\bar{w}_t)'\Psi \underline{w}_t + u_t^*$$

$$= \underline{\psi}'_1\ \underline{w}_t + \ (k(\bar{w}_t) \otimes\ \underline{w}_t\ )'\ \mathrm{vec}\ (\Psi) + u_t^* \qquad (3.6)$$

The test can be carried out as before, at the second stage y_t is regressed on w_t and $(k(\bar{w}_t) \otimes \underline{w}_t)$ and under $H'_{01} : \mathrm{vec}(\Psi) = 0$ the test statistic has an asymptotic $\chi^2\ (nr\ )$ distribution.

From (3.6) it is seen that the original null hypothesis H_{02} has been transformed into H'_{01}: vec $(\Psi) = 0$. Approximating f_1 as in (3.4) and reparameterizing the model may be seen as a way of removing the identification problem. However, it may also be seen as a solution in the spirit of Davies (1977). Let $\tilde{\underline{u}}^*$ be the residual vector from the regression (3.6). Then

$$\tilde{\underline{u}}^{*\prime} \tilde{\underline{u}}^* = \inf\ _{\theta_2\ ,\ \theta_3}\ \underline{u}(\underline{\theta})' \underline{u}(\underline{\theta})$$

and the test statistic

$$\bar{F} = \sup\ _{\theta_2\ ,\ \theta_3}\ F\ (\ \theta_2\ ,\ \theta_3\ ) = \frac{[\tilde{\underline{u}}' \tilde{\underline{u}} - inf\ \underline{u}\ (\underline{\theta}\ )' \underline{u}(\ \underline{\theta}\ )\ ]/nr}{inf.\ \underline{u}(\ \underline{\theta}\ )' \underline{u}(\ \underline{\theta}\ )/(T-n-nr\ )}\ .$$

The price of the neat asymptotic null distribution is that not all the information in Ψ has been used: in fact Ψ is of rank one and only contains n+r+1 parameters.

As an example, assume $\underline{w}_t = \bar{\underline{w}}_t = (y_{t-1},...,y_{t-p})'$, choose $\theta_{30} = 0$, and let θ_{31} be a scalar and $k(\bar{w}_t) = y_{t-1}^2$. This gives $\Psi = \theta_{31}\theta_2\theta_1$ and $(k(\bar{w}_t) \otimes \bar{w}_t) = (y_{t-1}^3, y_{t-1}^2 y_{t-2}\ ,...,\ y_{t-1}^2 y_{t-p})$. The resulting test is the linearity test against the univariate exponential autoregressive model in Saikkonen and Luukkonen (1988). If $\bar{\underline{w}}_t = k(\bar{w}_t)$, $\sum_{i=1}^{p} \sum_{j=i}^{p} \varphi_{ij} y_{t-i} y_{t-j}$ replaces $(k\ (\bar{w}_t) \otimes\ \underline{w}_t)'$ *vec* (Ψ) and $H'_{01} : \varphi_{ij} = 0,\ i = 1,...,p;\ j = i,...p$. The test is the first of the three linearity tests against smooth transition autoregression in Luukkonen et al. (1988b)

when the delay parameter d is unknown but it is assumed that $1 \leq d \leq p$. The number of degrees of freedom in the asymptotic null distribution equals $p(p+1)/2$. If w_t also contains other variables than lags of y_t, the test is a linearity test against smooth transition regression; see Granger and Teräsvirta (1992, chapter 6). If the delay parameter is known, $k(\underline{w}_t) = y_{t-d}$, so that $(k(\underline{w}_t) \otimes \underline{w}_t) = (y_{t-1}y_{t-d}, \ldots, y_{t-d}^2, \ldots, y_{t-p}y_{t-d})'$ and the F test has p and $T-n-p$ degrees of freedom.

In some cases the first-order Taylor series approximation is inadequate. For instance, let $\underline{\theta}_1 = (\theta_{10},0,\ldots,0)'$ in (3.3) so that the only nonlinearity is described by f_1 multiplied by a constant. Then the LM type test has no power against the alternative because $(k(\underline{w}_t) \otimes \underline{w}_t)'$ vec(ψ_1) $= \underline{\alpha}'\underline{w}_t$, say, and therefore $\varphi_{ij} = 0$, $\forall i,j$. In such a situation, a third-order Taylor series approximation of f is needed for constructing a proper test; see Luukkonen et al. (1988b) for discussion.

3.2. Tests without a specific alternative

The above linearity tests are tests against a well-specified nonlinear alternative. There exist other tests that are intended as general tests without a specific alternative. We shall consider some of them. The first one is the Regression Error Specification Test (RESET; Ramsey, 1969). Suppose we have a linear model

$$y_t = \underline{\varphi}'\underline{w}_t + u_t \tag{3.7}$$

where w_t is as in (3.1) and whose parameters we estimate by OLS. Let $\tilde{u}_t$, $t = 1,\ldots,T$, be the estimated residuals and $\tilde{y}_t = y_t - \tilde{u}_t$ the fitted values. Construct an auxiliary regression

$$\tilde{u}_t = \underline{\psi}'\underline{w}_t + \sum_{j=2}^{h} \delta_j \tilde{y}_t^j + u_t^* . \tag{3.8}$$

The RESET is the F-test of the hypothesis H_0: $\delta_j = 0$, $j = 2,\ldots,h$, in (3.8). If $\underline{w}_t = (1, y_{t-1},\ldots,y_{t-p})'$ and $h = 2$, (3.8) yields the univariate linearity test of Keenan (1985). In fact, RESET may also be interpreted as a LM test against a well-specified alternative; see for instance Teräsvirta (1990b) or Granger and Teräsvirta (1992, chapter 6).

Tsay (1986) suggested augmenting the univariate (3.7) by second-order terms so that the auxiliary regression corresponding to (3.8) becomes

$$\tilde{u}_t = \psi' w_t + \sum_{i=j}^{p} \sum_{j=i}^{p} \varphi_{ij} y_{t-i} y_{t-j} + u_t^* \tag{3.9}$$

The linearity hypothesis to be tested is H_0: $\varphi_{ij} = 0$, $\forall$i,j. The generalization to multivariate models is immediate. This test also has a LM type interpretation showing that the test has power against a larger variety of nonlinear models than the RESET. This is seen by comparing (3.9) with (3.6) when $k(\underline{\overline{w}}_t) = \underline{\overline{w}}_t$ as discussed in the previous section. The advantage of RESET lies in the small number of parameters in the null hypothesis. When $\underline{w}_t = (1, y_{t-1})'$ (or $\underline{w}_t = (1, x_{tj})'$), the two tests are identical.

A general linearity test can also be based on the neural network model (2.4), and such a test is presented in Lee et al. (1992). In computing the test statistic, γ_j, $j = 1,...,p$, in (2.4) are selected randomly from a distribution. Teräsvirta et al. (1991) showed that this can be avoided by deriving the test by applying the LM principle, in which case $p = 1$ in (2.4). Assuming $p > 1$ does not change anything because (2.4) is not globally identified under that assumption if $\varphi_j(x) = \varphi(x)$, $j = 1,...,p$. The auxiliary regression for the test becomes

$$\tilde{u}_t = \psi' w_t + \sum_{i=1}^{p} \sum_{j=i}^{p} \delta_{ij} y_{t-i} y_{t-j} + \sum_{i=1}^{p} \sum_{j=i}^{p} \sum_{k=j}^{p} \delta_{ijk} y_{t-i} y_{t-j} y_{t-k} + u_t^* \tag{3.10}$$

and the linearity hypothesis $H_0 : \delta_{ij} = 0$, $\delta_{ijk} = 0$, $\forall i,j,k$. The simulation results in Teräsvirta et al. (1991) indicate that in small samples the test based on (3.10) has better power than the original neural network test.

There has been no mention yet about tests against piecewise linear or switching regression or its univariate counterpart, threshold autoregression. The problem is that f_1 in (3.3) is not a continuous function of parameters if the switch-points or thresholds are unknown. This makes the likelihood function irregular and the score principle inapplicable. Ertel and Fowlkes (1976) suggested the use of cumulative sums of recursive residuals for testing linearity. First order the variables in ascending (or descending) order according to the transition variable. Compute the parameters recursively and consider the cumulative sum of the recursive residuals. The test is analogous to the CUSUM test Brown et al. (1975) suggested in which time is the transition variable and no lags of y_t are allowed in $\underline{w}_t$. However, Krämer et al. (1988) showed that the presence of lags of y_t in the model does not affect the asymptotic null distribution of the CUSUM statistic. Even before that, Petruccelli and Davies (1986) proposed the same test for the univariate (threshold autoregressive) case;

see also Petruccelli (1990). The CUSUM test may also be based on residuals from OLS estimation using all the observations instead of recursive residuals. Ploberger and Krämer (1992) recently discussed this possibility.

The CUSUM principle is not the only one available from the literature of structural change. Quandt (1960) suggested generalizing the F test (Chow, 1960) for testing parameter constancy in a linear model with known change-point by applying $\overline{F} = \sup_{t \in T} F(t)$ where $T = \{t | t_0 < t < T - t_1\}$. He noticed that the null distribution of $\overline{F}$ was nonstandard. Andrews (1990) provided the asymptotic null distribution for $\overline{F}$ and tables for critical values; see also Hansen (1990). If the observations are ordered according to a variable other than time, a linearity test against switching regression is obtained. In the univariate case, Chan (1990) and Chan and Tong (1990) applied the idea of Quandt to testing linearity against threshold autoregression (TAR) with a single threshold; see also Tong (1990, chapter 5). Chan (1991) provided tables of percentage points of the null distribution of the test statistic. In fact, this test can be regarded as one against a well-specified alternative: a two-regime switching regression or threshold autoregressive model with a known transition variable or delay parameter. For further discussion, see Granger and Teräsvirta (1992, chapter 6).

Petruccelli (1990) compared the small sample performance of the CUSUM, the threshold autoregression test of Chan and Tong and the LM type test against logistic STAR of Luukkonen et al. (1988b) when the true model was a single-threshold TAR model. The results showed that the first two tests performed reasonably well (as the CUSUM test a "reverse CUSUM" (Petruccelli, 1990) was used). However, they also demonstrated that the LM type test had quite comparable power against this TAR which is a special case of the logistic STAR model.

As mentioned in the introduction, Brock et al. (1987) proposed a test (BDS test) of independent, identically distributed observations based on the correlation integral, a concept that arises in chaos theory. Let $Y_{t,n}$ be a part of a time series $Y_{T,T} = (y_T,...,y_1)$: $Y_{t,n} = (y_t, y_{t-1},...,y_{t-n+1})$. Compare a pair of such vectors $Y_{t,n}$ and $Y_{s,n}$. They are said to be no more than ε apart if

$$\| Y_{t,j} - Y_{s,j} \| \leq \varepsilon \, , j = 0,1,...,n-1. \tag{3.11}$$

The correlation integral is defined as

$$C_n(\varepsilon) = \lim_{T \to \infty} T^{-2} \{\text{number of pairs } (t,s) \text{ with } 1 \leq t, s \leq T \text{ such that (3.11) holds}\} .$$

Brock et al. (1987) defined

$$S(n,\varepsilon) = \hat{C}_n(\varepsilon) - [\hat{C}_1(\varepsilon)]^n . \tag{3.12}$$

Under the hypothesis that $\{y_t\}$ is an iid process, (3.12) has an asymptotic normal distribution with zero mean and variance given in Brock et al. (1987). Note that (3.12) depends on n and ε which the investigator has to choose. A much more thorough discussion of the BDS test and its properties is found in Brock and Potter (1992) or Scheinkman (1990). It may be mentioned, however, that a rather long time series is needed to obtain reasonable power. Lee et al. (1992) contains some small-sample evidence on the behaviour of the BDS test but it is not very conclusive; see Teräsvirta (1990b).

Linearity of a single series may also be tested in the frequency domain. Let $\{y_t\}$ be stationary and have finite moments up to the sixth order. Then we can define the bispectral density $f(\omega_i,\omega_j)$ of y_t based on third moments and

$$b(\omega_i, \omega_j) = \frac{|f(\omega_i, \omega_j)|^2}{f(\omega_i)f(\omega_j)f(\omega_i + \omega_j)} .$$

where $f(\omega_i)$ is the spectral density of y_t. Two hypotheses can be tested: (i) if $f(\omega_i,\omega_j) \equiv 0$ then y_t is linear and Gaussian, (ii) if $b(\omega_i,\omega_j) \equiv b_0 > 0$ then y_t is linear but not Gaussian, i.e., the parameterized linear model for $\{y_t\}$ has non-Gaussian errors. Subba Rao and Gabr (1980) proposed tests for testing these two hypothesis. Hinich (1982) derived somewhat different tests for the same purpose. For more discussion see e.g. Priestley (1988) and Brockett et al. (1988). A disadvantage of these tests seems to be relatively low power in small samples. Besides, performing the tests requires more computation than carrying out most of their time domain counterparts.

It has been assumed so far that $g \equiv 1$ in (3.1). If this assumption is not satisfied, the size of the test may be affected. At least the BDS test and the tests based on bispectral density are known to be sensitive to departures from that assumption. If linearity of the conditional mean is tested against a well-specified alternative using LM type tests, some possibilities of taking conditional heteroskedasticity into account exist and will be briefly mentioned in the next section.

3.3. Constancy of conditional variance

The assumption $g \equiv 1$ is also a testable hypothesis. However, because conditional heteroskedasticity is discussed elsewhere in this volume, testing $g \equiv 1$ against nonconstant conditional variance is not considered here. This concerns not only testing linearity against ARCH but also testing it against random coefficient linear regression; see e.g. Nicholls and Pagan (1985) for further discussion on the latter situation.

If $f \equiv 0$ and $g \equiv 1$ are tested jointly, a typical LM or LM type test is a sum of two separate LM (type) tests for $f \equiv 0$ and $g \equiv 1$, respectively. This is the case because under this joint null hypothesis the information matrix is block diagonal; see Granger and Teräsvirta (1992, chapter 6). Higgins and Bera (1989) derived a joint LM test against bilinearity and ARCH. On the other hand, testing $f \equiv 0$ when $g \not\equiv 1$ is a more complicated affair than it is when $g \equiv 1$. If g is parameterized, the null model has to be estimated under conditional heteroskedasticity. Besides, it may no longer be possible to carry out the test making use of a simple auxiliary regression, see Granger and Teräsvirta (1992). If g is not parameterized but $g \not\equiv 1$ is suspected then the tests described in section 3.1 as well as RESET and the Tsay test can be made robust against $g \not\equiv 1$. Davidson and MacKinnon (1985) and Wooldridge (1990) described techniques for doing this. The present simulation evidence is not yet sufficient to fully evaluate their performance in small samples.

4. SPECIFICATION OF NONLINEAR MODELS

If linearity tests indicate the need for a nonlinear model and economic theory does not suggest a completely specificied model, then the structure of the model has to be specified from the data. This problem also exists in nonparametric modelling as a variable selection problem because the lags needed to describe the dynamics of the process are usually unknown; see Auestad and Tjøstheim (1991) and Tjøstheim and Auestad (1991a,b). To specify univariate time series models, Haggan et al. (1984) devised a specification technique based on recursive estimation of parameters of a linear autoregressive model. The parameters of the model were assumed to change over time in a certain fashion. Choosing a model from a class of state-dependent models, see Priestley (1980, 1988), was carried out by examining the graphs of recursive estimates. Perhaps because the family of state-dependent models is large and the possibilities thus many, the technique is not easy to apply.

If the class of parametric models to choose from is more restricted, more concrete specifi-

cation methods may be developed. (For instance, Box and Jenkins (1970) restricted their attention to linear ARMA models.) Tsay (1989) presented a technique making use of linearity tests and visual inspection of some graphs to specify a model from the class of threshold autoregressive models. It is easy to use and seems to work well. Chen and Tsay (1990) considered the specification of functional-coefficient autoregressive models whereas Chen and Tsay (1991) extended the discussion to additive functional coefficient regression models. The key element in that procedure is the use of arranged local regressions in which the observations are ordered according to a transition variable. Lewis and Stevens (1991a) applied multivariate adaptive regression splines (MARS), see Friedman (1991), to specify adaptive spline threshold autoregressive models. Teräsvirta (1990a) discussed the specification of smooth transition autoregressive models. This technique was generalized to smooth transition regression models in Granger and Teräsvirta (1992, chapter 7) and will be considered next.

Consider the smooth transition regression model with $p+k+1$ independent variables

$$y_t = \underline{\varphi}'\underline{w}_t + (\underline{\theta}'\underline{w}_t)\, F(z_t) + u_t \tag{4.1}$$

where $E\{u_t|I_t\} = 0$, $\text{cov}\{u_t|I\} = \sigma^2$, $I_t = \{y_{t-j},\ j=1,2,...;x_{t-j,i},\ i=1,...,k,\ j=0,1,...\}$, (information set), $\underline{\varphi} = (\varphi_0,\varphi_1,...,\varphi_m)'$, $\underline{\theta} = (\theta_0,\theta_1,...,\theta_m)'$, $m = k+p+1$, and $\underline{w}_t = (1,y_{t-1},...,y_{t-p};\ x_{1t},...,x_{kt})'$. The alternatives for F are $F(z_t) = (1 + \exp\{-\gamma(z_t - c)\})^{-1}$, $\gamma > 0$, which gives the logistic STR model and $F(z_t) = 1 - \exp\{-\gamma(z_t - c)^2\}$, $\gamma > 0$, corresponding to the exponential STR model. The transition variable z_t may be any element of w_t other than 1 or another variable not included in w_t.

The specification proceeds in three stages. First, specify a linear model to serve as a base for testing linearity. Second, test linearity against STR using the linear model as the null model. If linearity is rejected, determine the transition variable from the data. Testing linearity against STR is not difficult. A test with power against both LSTR and ESTR if the transition variable is assumed known is obtained by proceeding as in section 3.1. This leads to the auxiliary regression

$$\hat{u}_t = \underline{\beta}_0'\underline{w}_t + \underline{\beta}_1'\underline{w}_t\, z_{td} + \underline{\beta}_2'\underline{w}_t\, z_{td}^2 + \underline{\beta}_3'\underline{w}_t\, z_{td}^3 + \eta_t \tag{4.2}$$

where z_{td} is the transition variable and $\hat{u}_t$ is the OLS residual from the linear regression $y_t = \beta'w_t + u_t$. If z_{td} is an element of $\underline{w}_t$, $\underline{w}_t = (1, \underline{\tilde{w}}_t')'$ has to be replaced by $\underline{\tilde{w}}_t$ in (4.2). The linearity hypothesis is H_{0d}: $\underline{\beta}_1 = \underline{\beta}_2 = \underline{\beta}_3 = 0$. Equation (4.2) is also used for selecting z_{td}.

The test is carried out for all candidates for z_{td}, and the one yielding the smallest p-value is selected if that value is sufficiently small. If it is not, the model is taken to be linear. This procedure is motivated as follows. Suppose there is a true STR model with transition variable z_{td} that generated the data. Then the LM type test against that alternative has optimal power properties. If an inppropriate transition variable is selected for the test, the resulting test may still have power against the true alternative but the power is less than if the correct transition variable is used. Thus the strongest rejection of the null hypothesis suggests that the corresponding transition variable be selected. For more discussion of this procedure see Teräsvirta (1990a,c) and Granger and Teräsvirta (1992, chapter 6 and 7). If linearity is rejected and a transition variable selected, then the third step is to choose between LSTR and ESTR models. This can be made by testing a set of nested null hypotheses within (4.2): they are H^*_{03}: $\beta_3 = 0$, H^*_{02}: $\beta_2 = 0|\beta_3 = 0$ and H^*_{01}: $\beta_1 = 0|\beta_2 = \beta_3 = 0$. The test results contain information that is used in making the choice; see Granger and Teräsvirta (1992, chapter 7).

Specifying the lag structure of (4.1) could be done within (4.2) using an appropriate model selection criterion but there is little experience about the success of such a procedure. In the existing applications, a general-to-specific approach based on estimating nonlinear STR (or STAR) models has mostly been used.

The model specification problem also arises in nonparametric time series modelling. Taking model (2.5) as a starting-point, there is the question of which lags $x_{t-i_1}, \ldots, x_{t-i_p}; y_{t-j_1}, \ldots, y_{t-j_q}$ should be included in the model. Furthermore it should be investigated whether the functions f and g are linear or nonlinear and whether they are additive or not. Moreover, if interaction terms are included, how should they be modelled and, more generally, can the nonparametric analysis suggest functional forms such as the smooth transition or threshold function or an ARCH type function for conditional variance?

These are problems of exploratory data analysis for nonlinear time series, and relatively little nonparametric work has been done in the area. Various graphical model indicators have been tried out in Tong (1990, chapter 7), Haggan et al. (1984) and Auestad and Tjøstheim (1990), however. Perhaps the most natural quantities to look at are the lagged conditional mean and variance of increasing order, i.e.

$$M_{y,k}(y) = E(y_t \mid y_{t-k} = y) \qquad M_{y,k}(x) = E(y_t \mid x_{t-k} = x)$$

$$V_{y,k}(y) = var(y_t \mid y_{t-k} = y) \qquad V_{y,k}(x) = var(y_t \mid x_{t-k} = x) \tag{4.3}$$

In univariate modelling these quantities have been extensively, albeit informally, used in Tong (1990, chapter 7). They can give a rough idea of the type of nonlinearity involved, but they fail to reveal things like the lag structure of an additive model.

A more precise and obvious alternative is to look at the functions $M(\underline{y};\underline{x})$ and $V(\underline{y};\underline{x})$ defined in (2.6) and (2.7), but they cannot be graphically displayed for $p+q > 2$, and the curse of dimensionality quickly becomes a severe problem. Auestad and Tjøstheim (1991) and Tjøstheim and Auestad (1991a) introduced projections as a compromise between $M(\underline{y};\underline{x})$, $V(\underline{y};\underline{x})$ and the indicators (4.3). To define projections consider the conditional mean function $M(y_{t-i_1},\ldots, y_k,\ldots, y_{t-i_p}; x_{t-j_1},\ldots, x_{t-j_q})$ with y_{t-i_k} excluded. The one-dimensional projector of order (p,q) projecting on lag i_k of y_t is defined by

$$P_{y,k}(y) = E\left\{ M(y_{t-i_1},\ldots, y,\ldots, y_{t-i_p}; x_{t-j_1},\ldots, x_{t-j_q}) \right\}. \tag{4.4}$$

The projector $P_{x,k}(x)$ is defined in the same way. For an additive model with $M(y_1,\ldots, y_p; x_1,\ldots, x_q) = \sum_{i=1}^{p} \alpha_i(y_i) + \sum_{j=1}^{q} \beta_j(x_j)$ it is easily seen that if all p+q lags are included in the projection operation, then

$$P_{y,k}(y) = \alpha_k(y) + \mu_k \qquad P_{x,k}(x) = \beta_k(x) + \theta_k$$

where $\mu_k = E(y_t) - E\{\alpha_k(y_t)\}$ and $\theta_k = E(x_t) - E\{\beta_k(x_t)\}$. Clearly the additive terms $\alpha_k(y)$ and $\beta_k(x)$ cannot be recovered using $M_{y,k}$ and $M_{x,k}$ of (4.3).

Projectors can be defined similarly for the conditional variance, and in principle they reveal the structure of models having an additive conditional variance function. Both types of projectors can be estimated by replacing theoretical expectations with empirical averages and by introducing a weight function to screen off extreme data. Properties and details are given in Auestad and Tjøstheim (1991) and Tjøstheim and Auestad (1991a).

An important part of the model specification problem consists of singling out the significant lags $i_1,\ldots,i_p; j_1,\ldots,j_q$ and the orders p and q for the conditional mean (2.6) and conditional variance (2.7). Auestad and Tjøstheim (1990, 1991), Tjøstheim and Auestad (1991b) and Cheng and Tong (1990) considered this problem, Granger and Lin (1991) did the same from a somewhat different point of view. Auestad and Tjøstheim adopted an approach analogous to the parametric final prediction error (FPE) criterion of Akaike (1969). They treated it

only in the univariate case, but it is easily extended to the multivariate situation.

For model (2.5) with $g \equiv 1$ (no heterogeneity) a nonparametric bias corrected and estimated version of the parametric FPE is given by

$$\hat{FPE}\,(i_1,\ldots,i_r\,;j_1,\ldots,j_s)=$$

$$T^{-1}\sum_t (y_t-\hat{M}\,(y_{t-i_1},\ldots,x_{t-i_r}\,;x_{t-j_1},\ldots,x_{t-j_s})\,\frac{1+(nh^{r+s})^{-1}J^{r+s}B_{r,s}}{1-(nh^{r+s})^{-1}\left\{2k(0)^{r+s}-J^{r+s}\right\}B_{r,s}} \tag{4.5}$$

where $J=\int k^2(x)dx$, k is a kernel function and $B_{r,s}$ essentially represents the dynamic range of the data in an (r+s)-dimensional space. It can be estimated as

$$B_{r,s}=T^{-1}\sum_t \frac{w^2\,(y_{t-i_1},\ldots,y_{t-i_r}\,;x_{t-j_1},\ldots,x_{t-j_s})}{\hat{p}\,(y_{t-i_1},\ldots,y_{t-i_r}\,;x_{t-j_1},\ldots,x_{t-j_s})}$$

where w is a weight function designed to screen off extreme observations, and $\hat{p}$ is the estimated joint density function. Criterion (4.5) can be used both to determine orders p and q and significant lags $i_1,..,i_p$ and $j_1,...,j_q$ in (2.6). A more general formula for the heterogeneous case is given in Tjøstheim and Auestad (1991b) to which the reader is referred for details of derivation and examples with simulated and real data. Cheng and Tong (1992) discussed a closely related approach based on cross validation.

An alternative and less computer intensive method is outlined by Granger and Lin (1991). They use the Kendall rank partial autocorrelation function and the bivariate information measure

$$\int \frac{\log \hat{p}\,(x,y)}{\log\,(\hat{p}\,(x)\,\hat{p}\,(y))}\,\hat{p}\,(x,y)dxdy$$

for a pair of lags. Joe (1989) studied its properties in the iid case. Robinson (1991) considered the random process case and tests of independence. Specification of semiparametric time series models is discussed in the next section together with estimation.

5. ESTIMATION IN NONLINEAR TIME SERIES

5.1. Estimation of parameters in parametric models

For parametric nonlinear models, conditional nonlinear least squares is the most common estimation technique. If the errors are normal and independent, this is equivalent to conditional maximum likelihood. The theory derived for dynamic nonlinear models (3.1) with $g \equiv 1$ gives the conditions for consistency and asymptotic normality of the estimators. For an account, see e.g. Gallant (1987, chapter 7). Even more general conditions were recently laid out in Pötscher and Prucha (1990, 1991). These conditions may be difficult to verify in practice, so that the asymptotic standard deviation estimates, confidence intervals and the like have to be interpreted with care. For discussions of estimation algorithms see e.g. Quandt (1983), Judge et al. (1985, appendix B) and Bates and Watts (1988). The estimation of parameters in (2.2) may not always be straightforward. Local minima may occur, so that estimation with different starting-values is recommended. Estimation of γ in transition function (2.3) may create problems if the transition is rapid because there may not be sufficiently many observations in the neighbourhood of the point about which the transition takes place. The convergence of the estimate sequence may therefore be slow, see Bates and Watts (1988, p. 87) and Granger and Teräsvirta (1992, chapter 7). For simulation evidence and estimation using real economic data sets see also Granger et al. (1992), Luukkonen (1990), Teräsvirta (1990a) and Teräsvirta and Anderson (1991). Model (2.2) may even be a switching regression model in which case γ is not finite and cannot be estimated. In that case its estimated value will grow until the iterative estimation algorithm breaks down. An available alternative is then to fix γ at some sufficiently large value and estimate the remaining parameters conditionally on that value.

The estimation of parameters becomes more complicated if the model contains lagged errors as the bilinear model does. Subba Rao and Gabr (1984) outlined a procedure for the estimation of a bilinear model based on maximizing the conditional likelihood. Quick preliminary estimates may be obtained using a long autoregression to estimate the residuals and OLS for estimating the parameters keeping the residuals fixed. This is possible because the bilinear model has a simple structure in the sense that it is linear in the parameters if we regard the lagged residuals as observed. Granger and Teräsvirta (1992, chapter 7) suggested this alternative.

If the model is a switching regression or threshold autoregressive model, nonlinear least squares is an inapplicable technique because of the irregularity of the sum of squares or likelihood function. The problem consists of the unknown switch-points or thresholds for

which unique point estimates are not available as long as the number of observations is finite. Tsay (1989) suggested specifying (approximate) switch-points from "scatterplots of t-values" in ordered (according to the switching variable) recursive regressions. As long as the recursion stays in the same regime, the *t*-value of a coefficient estimate converges to a fixed value. When observations from another regime are added into the regression, the coefficient estimates start changing and the *t*-values deviating. Tsay (1989) contains examples. The estimation of parameters in regimes is carried out by ordinary least squares. Chan (1988) showed (in the univariate case) that if the model is stationary and ergodic, the parameter estimates, including those of the thresholds, are strongly consistent.

5.2. Estimation of nonparametric functions

In nonparametric estimation the most common way of estimating the conditional mean (2.6) and variance (2.7) is to apply the so-called kernel method. It is based on a kernel function $k(x)$ which typically is a real continuous, bounded, symmetric function integrating to one. Usually it is required that $k(x) \geq 0$ for all x, but sometimes it is advantageous to allow $k(x)$ to take negative values, so that we may have $\int x^2 k(x)dx = 0$. The kernel method is explained in much greater detail in the chapter by Härdle on nonparametric estimation.

The kernel acts as a smoothing device in the estimation procedure. For quantities depending on several variables as in (2.6) and (2.7) a product kernel can be used. Then the kernel estimates of M and V are

$$\hat{M}(y_1,\ldots,y_p, x_1,\ldots,x_q) = \frac{\frac{1}{T}\sum_s y_s \prod_{r=1}^{p} k_{h,1}(y_r - y_{s-i_r}) \prod_{r=1}^{q} k_{h,2}(x_r - x_{s-i_r})}{\frac{1}{T}\sum_s \prod_{r=1}^{p} k_{h,1}(y_r - y_{s-i_r}) \prod_{r=1}^{q} k_{h,2}(x_r - x_{s-i_r})} \tag{5.1}$$

$$\hat{V}(y_1,\ldots,y_p, x_1,\ldots,x_q) = \frac{\frac{1}{T}\sum_s y_s^2 \prod_{r=1}^{p} k_{h,1}(y_r - y_{s-i_r}) \prod_{r=1}^{q} k_{h,2}(x_r - x_{s-i_r})}{\frac{1}{T}\sum_s \prod_{r=1}^{p} k_{h,1}(y_r - y_{s-i_r}) \prod_{r=1}^{q} k_{h,2}(x_r - x_{s-i_r})} - (\hat{M}(\underline{y}, \underline{x}))^2 \tag{5.2}$$

where $k_{h,i}(x) = h_i^{-1} k_i(h_i^{-1}x)$, i=1,2. Here k_1 and k_2 are the kernel functions associated with the $\{y_t\}$ and $\{x_t\}$ process, and h_1 and h_2 are the corresponding bandwidths. The bandwidth

controls the width of the kernel function and thus the amount of smoothing involved. The bandwidth will depend on the total number of observations T, so that $h = h(T) \to 0$ as $T \to \infty$. It also depends on the dimensions p and q, but this has been suppressed in the above notation. In the following, to simplify notation, it is assumed that $\{y_t\}$, $\{x_t\}$ are measured roughly on the same scale, so that the same bandwidth and the same kernel function can be used everywhere.

Under regularity conditions (Robinson, 1983) it can be proved that $\hat{M}(\underline{y}, \underline{x})$ and $\hat{V}(\underline{y}, \underline{x})$ are asymptotically normal. More precisely,

$$(Th^{p+q})^{1/2}\{\hat{M}(\underline{y}, \underline{x}) - M(\underline{y}, \underline{x})\} \to N(0, \frac{V(\underline{y}, \underline{x})}{p(\underline{y}, \underline{x})} J^{p+q}) \tag{5.3}$$

and

$$(Th^{p+q})^{1/2}\{\hat{V}(\underline{y}, \underline{x}) - V(\underline{y}, \underline{x})\} \to N(0, \frac{s(\underline{y}, \underline{x})}{p(\underline{y}, \underline{x})} J^{p+q}) \tag{5.4}$$

where the convergence is in distribution, $J = \int k^2(x)dx$, and $s(\underline{y}, \underline{x})$ is defined in Auestad and Tjøstheim (1990).

Several points should be noted for (5.3) and (5.4). For parametric models we have $\sqrt{T}$-consistency. For nonparametric models the rate is $\sqrt{Th^{p+q}}$, which is slower. The presence of $p(\underline{y},\underline{x})$ in the denominator of the left-hand sides of (5.3) and (5.4) means that the variance blows up close to the boundaries of the data set, and extreme care must be used in the interpretation of $\hat{M}(\underline{x}, \underline{y})$ and $\hat{V}(\underline{x}, \underline{y})$ there.

There are other aspects of practical significance that are not immediately transparent from (5.3) and (5.4). They will be discussed next.

Confidence intervals. Asymptotic confidence intervals can in principle be computed from (5.3) and (5.4) by replacing $p(\underline{y},\underline{x})$, $V(\underline{y},\underline{x})$ and $s(\underline{y},\underline{x})$ by corresponding estimated quantities. An alternative is to try to form bootstrap confidence intervals. Franke and Wendel (1990) discussed a simple example where the bootstrap performs much better than asymptotic intervals. In the general case the bootstrap developed by Künsch (1989) and Politis and Romano (1990) may be needed.

Bias. As seen from (5.3) and (5.4), $\hat{M}(y, x)$ and $\hat{V}(y, x)$ are asymptotically unbiased. For a

finite sample size the bias can be substantial. Thus, reasoning as in Auestad and Tjøstheim (1990) yields

$$E\{\hat{M}(\underline{y},\underline{x})\} \sim M(\underline{y},\underline{x}) + h^2 I_2 p^{-1}(\underline{y},\underline{x})\{\sum_{i=1}^{p} \frac{\partial M(\underline{y},\underline{x})}{\partial y_i} \frac{\partial p(\underline{y},\underline{x})}{\partial y_i} +$$

$$\sum_{i=1}^{q} \frac{\partial M(\underline{y},\underline{x})}{\partial x_i} \frac{\partial p(\underline{y},\underline{x})}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{p} \frac{\partial^2 M(\underline{y},\underline{x})}{\partial y_i^2} p(\underline{y},\underline{x}) + \frac{1}{2}\sum_{i=1}^{q} \frac{\partial^2 M(\underline{y},\underline{x})}{\partial x_i^2} p(\underline{y},\underline{x})\} \tag{5.5}$$

where $I_2 = \int x^2 k(x)dx$. A corresponding formula (Tjøstheim and Auestad, 1991a) holds for the conditional variance. A Gaussian linear model will have a linear bias in the conditional mean, but in general the bias can lead to a misspecified model. For example a model with a flat conditional variance (no conditional heteroskedasticity) may in fact appear to have some form of heteroskedasticity due to bias from a rapidly varying $M(\underline{y},\underline{x})$. An example is given in Auestad and Tjøstheim (1990). Generally, $\hat{V}(\underline{y},\underline{x})$ is more affected by bias and has more variability than $\hat{M}(\underline{y},\underline{x})$. This makes it harder to reveal the structure of the conditional variance using purely nonparametric means; see for instance the example of conditional stock volatility in Pagan and Schwert (1990). Another problem is that misspecification of the conditional mean may mix up conditional mean and variance effects. This is of course a problem in parametric models as well.

Choosing the bandwidth: Comparing the variance and bias formulae (5.3-5) it is seen that the classical problem of all smoothing operations is present. As h increases, the variance decreases whereas the bias increases and vice versa. How should h be chosen for a given data set?

There are at least three approaches to this problem. The simplest solution is to compute estimates for several values of h and select one subjectively. A second possibility is to use asymptotic theory. From (5.3-5) it is seen that if we require that variance and bias squared should be asymptotically balanced, then $(Th^{p+q})^{-1} \sim h^4$, or $h \sim T^{-1/(p+q+4)}$. An extension of this argument (Truong and Stone, 1992) yields $h \sim T^{-1/(p+q+2R)}$, where R is a smoothness parameter. The problem of choosing the proportionality factor still remains. A discussion of this and related problems is given in Härdle (1990, chapter 5), in the chapter by Härdle in this volume and in Marron (1989). The third possibility, which is the most time consuming but possibly the one most used in practice, is to use some form of cross validation. For details, see the above references. Simulation experiments showing considerable variability

for h selected by cross validation for one and the same model have been reported.

Boundary effects. For a point $(\underline{y},\underline{x})$ close to the boundary of the data set there will be disproportionally more points on the "inward" side of $(\underline{y}, \underline{x})$. This asymmetry implies that we are not able to integrate over the entire support of the kernel function, so that we cannot exploit the fact that $\int xk(x)dx = 0$. This in turn means that there is an additional bias of order h due to this boundary effect. For example, for a linear regression model the estimated regression line would bend close to the boundary. The phenomenon has primarily been examined theoretically in the fixed regression design case (Rice, 1984; Müller, 1990).

Higher order kernels. Sometimes so-called higher order kernels have been suggested for reducing bias. It is seen from (5.4) that if k is chosen such that $\int x^2k(x)dx = 0$, the bias will effectively be reduced to the next order term in the bias expansion (typically of order h^4). However, practical experience in the finite sample case has been mixed, and a higher order kernel does not work unless T is rather large.

Curse of dimensionality. This problem was mentioned in the Introduction. It is a well-known difficulty of multidimensional data analysis and a serious one in nonparametric estimation. Although the bandwidth h typically increases somewhat as the dimensions p and q increase, this is by no means enough to compensate for the sparsity of points in a neighborhood of a given point. The estimate $\hat{M}(y,x)$ will eventually reach a limiting value essentially independent of the chosen h. Indeed, in the limiting situation the sums in the numerator and denominator of (5.1) and (5.2) will be completely dominated by the observational pair (y_s,x_s) closest to $(\underline{y},\underline{x})$ and in the limit $\hat{M}(x,y)$ becomes

$$M^*(\underline{y},\underline{x}) = y_{s^*(\underline{y},\underline{x})+1}$$

where $s^*(y,x)$ is the s for which $\| \underline{y}_s - y \|^2 + \| \underline{x}_s - x \|^2$ is minimized. There may still be some useful information left in $M^*(\underline{y},\underline{x})$ that can be used for specification purposes (Tjøstheim and Auestad, 1991a,b) or as input to iterative algorithms described in the next section, but it is of little use as an accurate estimate of $M(\underline{y},\underline{x})$.

In general one should try to avoid the curse of dimension by not looking at too many regressors simultaneously; i.e. by considering (2.6) and (2.7) such that while i_p and i_q may be large, p and q are not. This requires a method for singling out significant lags nonparametrically, which was discussed in section 4. Alternatively, the problem may be handled by applying more restricted models which will be considered in the next section.

Other estimation methods. There are a number of alternative nonparametric estimation methods. These are described in Härdle (1990, chapter 3) and Hastie and Tibshirani (1990, chapter 2). The most commonly used are spline smoothing, nearest neihgbour estimation, orthogonal series expansion and the regressogram. For all of these methods there is a smoothing parameter that must be chosen in analogy with the choice of bandwidth for the kernel smoother. The asymptotic properties of the resulting estimators are roughly similar to those in kernel estimation. The spline smoother (Silverman, 1984) can be rephrased asymptotically as a kernel estimator with negative sidelobes. Diebolt (1990) applied the regressogram to test for nonlinearity. Yakowitz (1987) considered nearest neighbour methods in time series. Further applications will be mentioned in the next section.

5.3. Estimation in restricted nonparametric and semiparametric models

As mentioned above, general nonparametric estimation with many variables leads to increased variability and problems with the curse of dimensionality. To alleviate these problems one can look at more restrictive models requiring particular forms for f and g in (2.5) or one can consider semi-parametric models. This section is devoted to models of that kind.

Additive models: Virtually all restrictive models have some sort of additivity built into them. In the simplest case (using consecutive lags)

$$y_t = \sum_{i=1}^{p} \alpha_i (y_{t-i}) + \sum_{i=1}^{q} \beta_i (x_{t-i}) + \varepsilon_t .$$

Regression versions of such models and generalizations with interaction terms are analysed extensively in Hastie and Tibshirani (1990) and references therein. Taking conditional expectations with respect to y_{t-i} and x_{t-j} simple identities are obtained which can be used as a basis for an iterative algorithm for computing the unknown functions α_i and β_j. The algorithm needs initial values of these functions. One possibility is to use either projections or simply a linear model for this purpose. Some examples and theoretical properties in the pure regression case are given by Hastie and Tibshirani. See also Chen and Tsay (1991).

The ACE algorithm treats a situation in which the dependent variable may be transformed as well, so that

$$h(y_t)=\sum_i \alpha_i (y_{t-i}) + \sum_i \beta_i (x_{t-i}) + \varepsilon_t .$$

The algorithm is perhaps test suited for a situation where $\alpha_i = 0$ for all i, so that there is a clear distinction between the input and output variables. The method was developed in Breiman and Friedman (1985). Some curious aspects of the ACE algorithm are highlighted in Hastie and Tibshirani (1990, p. 184-186). In view of the above comments it is perhaps not surprising that in a time series example Hallman (1990) obtained better results by using a version of backfitting (Tibshirani, 1988) than with the ACE algorithm.

Chen and Tsay (1990) considered a univariate model allowing certain interactions. Their functional coefficient autoregressive (FCAR) model is given as

$$y_t = f_1 (y_{t-i_1}, \ldots, y_{t-i_k})\, y_{t-1} + \ldots + f_p (y_{t-i_1}, \ldots, y_{t-i_k})\, y_{t-p} + \varepsilon_t$$

with $i_k \leq p$. By ordering the observations according to some variable or a known combination of them to an "ordered" local regression the authors proposed an iterative procedure for evaluating $f_1,\ldots,f_p$ and gave some theoretical properties. The procedure simplifies dramatically if all the f_j are one-dimensional. The authors fitted an FCAR model of this type to the chicken pox data of Sughihara et al. (1990). The fitted model seemed to point at a threshold autoregressive model. The forecasts from such a model subsequently fitted to the data had a MSE at least 30 % smaller than a seasonal ARMA model used as a comparison for forecasting 4-11 months ahead.

Projection pursuit type models. In our notation these models can be written as

$$y_t = \sum_{j=1}^{r} \beta_j (\underline{\gamma}_j' \underline{y}_{t-1} + \underline{\varphi}_j' \underline{x}_{t-1}) + e_t$$

where β_j, $j=1,\ldots,r$, are unknown functions, $\underline{\gamma}_j$ and $\underline{\varphi}_j$ are unknown vectors determining the direction of the j-th projector, and $\underline{y}_{t-1}$, $\underline{x}_{t-1}$ are as in (2.5). An iterative procedure (Friedman and Stuetzle, 1981) exists for deriving optimal projectors (projection pursuit step) and functions β_j. The curse of dimensionality is avoided since in the smoothing part of the algorithm it is exploited that β_j is a function of one scalar variable. For time series data, experience with this method is limited. A small simulation study Granger and Teräsvirta (1991) conducted gave marginal improvements compared to linear model fitting for the particular nonlinear models they considered. Projection pursuit models are related to neural network models, but for the latter the functions β_j are assumed known and often $\beta_j = \beta$, $j = 1,\ldots,r$, thus giving a parametric model class. The fitting of neural network models is

discussed in White (1989).

Regression trees, splines and MARS. Assume a model of form

$$y_t = f(\underline{y}_{t-1}, \underline{x}_t) + e_t$$

and approximate $f(\underline{y},\underline{x})$ in terms of simple basis functions $B_j(\underline{y},\underline{x})$ so that $f_{appr}(\underline{y},\underline{x}) = \sum_j c_j B_j(\underline{y}, \underline{x})$. In the regression tree approach (Breiman et al., 1984) f_{appr} is built up recursively from indicator functions $B_j(\underline{y},\underline{x}) = I\{(\underline{y},\underline{x}) \in R_j\}$ and the regions R_j are partitioned in the next step of the algorithm according to a certain pattern. As can be expected there are problems in fitting simple smooth functions like the linear model.

Friedman (1991) in his MARS (Multivariate Adaptive Regression Splines) methodology has made at least two important new contributions. First, to overcome the difficulty in fitting simple smooth functions Friedman proposed not to automatically eliminate the parent region R_j in the above recursive scheme for creating subregions. In subsequent iteration both the parent region and its corresponding subregions are eligible for further partitioning. This allows for much greater flexibility. The second contribution is to replace step functions by products of linear left and right truncated regression splines. The products make it possible to include interaction terms. For a detailed discussion the reader is referred to Friedman (1991).

Lewis and Stevens (1991a) applied MARS to time series, both simulated and real data. As for most of the techniques discussed in this section a number of input parameters are needed. Lewis and Stevens recommended running the model for several sets of parameters and then selecting a final model based on various specification/fitting tests. They fitted a model to the sunspot data which has 3 one-way, 3 two-way and 7 three way interaction terms. The MARS model produced better overall forecasts of the sunspot activity than the models applied before. In Lewis and Stevens (1991b) riverflow is fitted against temperature and precipitation and good results obtained. There are as yet no applications to economic data.

The MARS technology appears very promising but must of course be tested more extensively on real and simulated data sets. No asymptotic theory with confidence intervals is available yet.

Stepwise series expansion of conditional densities. In a sense the conditional density $p(y_t \mid \underline{y}_{t-1}, \underline{x}_{t-1})$ is the most natural quantity to look at in a joint modelling of $\{y_t, x_t\}$ since predictive distributions as well as the conditional mean and variance can all be derived from this

quantity. Gallant and Tauchen (1990) used this fact as their starting point.

The conditional density is estimated, to avoid the curse of dimensionality, by expanding it in Hermite polynomials. These are centred and scaled so that the conditional mean $M(\underline{y},\underline{x})$ and variance $V(\underline{y},\underline{x})$ play a prominent role. As a first approximation they are supposed to be linear Gaussian and of ARCH type, respectively.

Gallant et al. (1990) looked at econometric applications, notably to stock market data. In particular, they investigated the relationship between volatility of stock prices and volume. A main finding was that an asymmetry in the volatility of prices when studied by itself more or less disappears when volume is included as an additional conditional variable. Possible asymmetry in the conditional variance function (univariate case) has recently been studied by a number of investigators using both parametric and nonparametric methods; see Engle and Ng (1991) and references therein.

Semiparametric models. Another way of trying to eliminate the difficulties in evaluating high-dimensional conditional quantities is to assume nonlinear and nonparametric dependence in some of the predictors and parametric and usually linear dependence in others. An illustrative example is given by Engle et al. (1986) who modelled electricity sales using a number of predictor variables. It is natural to assume the impact of temperature on electricity consumption to be nonlinear, as both high and low temperatures lead to increased consumption, whereas a linear relationship may be assumed for the other regressors. A similar situation arose in Shumway et al. (1988) which is a study of mortality as a function of weather and pollution variables in the Los Angeles region.

In the context of model (2.5) with a linear dependence on lags of y_t and nonlinearity with respect to the exogenous variable $\{x_t\}$, we have

$$y_t = \underline{a}'\underline{y}_{t-1} + f(\underline{x}_{t-1}) + \varepsilon_t.$$

The modelling technique would depend somewhat on the dimension of $\underline{x}_{t-1}$. In the case where the argument of f is scalar, it can be incorporated in the backfitting algorithm of Hastie and Tibshirani (1990, p. 118). Under quite general assumptions it is possible to obtain $\sqrt{T}$-consistency for the parametric part as demonstrated by Heckman (1986) and Robinson (1988). Powell et al. (1989) developed the theory further and gave econometric applications.

6. EVALUATION OF ESTIMATED MODELS

After estimating a nonlinear time series model it is necessary to evaluate its properties to see if the specified and estimated model may be regarded as an adequate description of the relationship it was constructed to characterize. The residuals of the model can be subjected to various tests such as those against ARCH and normality. At least in the parametric case linearity of the time series was tested, and the same tests may now be performed on the residuals to see if the model adequately characterizes the nonlinearity the tests previously suggested. Note, however, that the asymptotic distribution of the Ljung-Box test statistic of no autocorrelation based on estimated residuals is not available, as the correct number of degrees of freedom is known only for the linear ARMA case. However, considering residual autocorrelations as such is informative. One should also study the stability of the model, which generally can only be done numerically by simulating the model without noise. The exogenous variables should be set on a constant level, for instance to equal their sample means. If the solution path diverges, the model should be rejected and respecification attempted. Other examples of a solution are a limit cycle or a stable singular point. See e.g. Ozaki (1985) for further discussion.

The out-of-sample prediction of the model is an important part of the evaluation process. The precision of the forecasts should be compared to those from the corresponding linear model. However, as mentioned in the Introduction, the results also depend on the data during the forecasting period. If there are no observations in the range in which nonlinearity of the model makes an impact, then the forecasts cannot be expected to be more accurate than those from a linear model. The check is thus negative: if the forecasts from the nonlinear model are significantly less accurate than those from the corresponding linear one, then the nonlinear specification should be reconsidered.

7. EXAMPLE

As a parametric example of the specification, estimation and evaluation cycle we shall consider the seasonally unadjusted logarithmic U.S. industrial output 1960(1) to 1986(4). This is one of the series analyzed in Teräsvirta and Anderson (1991). The four quarter differences (growth rate) contain strong fluctuations, and the problem is to find an adequate description of the series. Selecting the linear autoregressive model using AIC yields an AR(6) model. Results of the linearity tests against STAR when the delay, d, is varied from 1 to 9 are given in Table 1. The test is based on an auxiliary regression like (4.2) which is

Table 1. The p-values of the linearity test against STAR for delays d = 1,...,9, for the four-quarter differences of the logarithmic U.S. industrial output, 1960(1)-1986(4). The base model is AR(6).

d	1	2	3	4	5	6	7	8	9
p-value	0.22	0.22	0.0062	0.0086	0.060	0.35	0.65	0.46	0.55

$$
\begin{aligned}
y_t = {} & \underset{(0.0072)}{-0.021} + \underset{(0.12)}{0.35}y_{t-1} + \underset{(0.20)}{0.24}y_{t-3} - \underset{(0.19)}{1.03}y_{t-4} + \underset{(0.11)}{0.33}y_{t-9} \\
& + (\underset{(0.0072)}{0.021} + \underset{(0.15)}{1.16}y_{t-1} - \underset{(0.10)}{0.57}y_{t-2} - \underset{(0.20)}{0.24}y_{t-3} + \underset{(0.19)}{1.03}y_{t-4} - \underset{(0.11)}{0.33}y_{t-9}) \\
& \times (1 + \exp[\underset{(37)}{-49}\times \underset{(0.0007)}{17.5}(y_{t-3}-0.0061)])^{-1} + \hat{u}_t
\end{aligned}
\tag{7.1}
$$

univariate and the transition variable is y_{t-d}. It is seen from Table 1 that d=3 and d=4 are the two obvious possibilities. Both were tried, and results finally reported are based on d=3. The next step was to choose between ESTAR and LSTAR by testing H^*_{03}, H^*_{02} and H^*_{01}, respectively, as discussed in section 4. The choice was an LSTAR model. Specification of the dynamic structure of the model was carried out by estimating LSTAR models from general to specific. The final model is s=0.0176 (residual standard deviation), LB(12) = 7.5 (Ljung-Box test with 12 autocorrelations), ML(4) = 7.3 (0.12) (McLeod-Li test against ARCH; p-value in parentheses), JB = 16.8 (0.0002) (Jarque-Bera test against non-normality). Note that there are several exclusion restrictions of type $\varphi_j = -\theta_j$ in (7.1) that have been imposed as they do not contradict the data. The normality test indicates that there are outliers among the residuals. They are in fact large (in absolute value) negative residuals which can be interpreted as negative shocks to the system. Carrying out the linearity test for the residuals setting d = 3 yields p-value = 0.040 indicating that there may still be some nonlinearity left not captured by (7.1). The number of lags in (7.1) is fairly large mainly because the series is seasonally unadjusted.

A detail in the estimated equation worth pointing out separately is the large standard

deviation of γ. In order to obtain an idea of the size of γ the exponent of the transition function is standardized by multiplying it by the inverse of the estimated standard deviation of y_t which equals 17.5. Seen against that background $\hat{\gamma} = 49$ is a large number indicating that the transition function changes from zero to unity or vice versa very rapidly when y_{t-3} crosses 0.0061. However, the uncertainty of $\hat{\gamma}$ seems large. This is the situation described in Bates and Watts (1988, p. 87). The large standard deviation reflects the fact that a wide range of values of γ around 49 would give a very similar transition function. It is not due to linearity of the series. Many more observations in the neighbourhood of c would be needed to estimate γ accurately. As a matter of fact, the residual variance of (7.1) is less than two-thirds of that of the AR(6) model, which is a large reduction.

The long-term solution of (7.1) is interesting. The realizations starting from a given set of starting values display cycles of varying length but no regular limit cycle seems to appear. If the starting-values are changed slightly, another cyclical realization emerges which gradually drifts apart from the previous one. Thus any change in the starting-values affects the long-run solution. It is also instructive to study the characteristic polynomials of (7.1) at different values of the logistic transition function F (see (4.1)), zero and unity being most interesting as the extreme values. When $F = 0$, which corresponds to the recession, the equation has two pairs of complex roots with moduli 1.11 and 1.01, respectively. The roots of the expansion regime ($F = 1$) are a stationary complex pair with modulus 0.75. Thus the output recovers from a recession swiftly as explosive roots are needed to characterize that. On the other hand, there is little else in this characterization than large negative shocks pushing the industry from an expansion into a recession.

One-quarter-ahead forecasts for 1987-1988 not shown here contribute little to the analysis because the output growth fluctuates little during that period. Both linear and nonlinear models have the root mean square prediction error way below the residual standard deviation of the nonlinear model.

The general conclusion is that there is no inherent nonlinearity in the series. For instance, it cannot be argued that the series is asymmetric, which has been a recent topic of nonlinear time series analysis. Nonlinear structure is needed mainly to describe the recovery of industrial production from a large negative shock. This is also seen by comparing the residuals of (7.1) and the AR(6) model, which also is a useful way of evaluating the nonlinear model. For more discussion of this and similar models for a few other OECD countries the reader is referred to Teräsvirta and Anderson (1991). Applications of bivariate STR models can be found in Granger et al. (1992); see also Granger and Teräsvirta (1992, chapter 10).

8. CONCLUSIONS

This chapter is an attempt at an overview of various ways of modelling nonlinear economic relationships. Since nonlinear time series models and methods are a very large field, not all important developments have been covered. The emphasis has been on model building, and the modelling cycle of linearity testing, model specification, parameter or nonparametric function estimation and model evaluation has been highlighted. The estimation of fully specified nonlinear theory models like disequilibrium models has not been the topic here. A majority of results concern the estimation of the conditional mean of a process whereas the conditional variance has received less attention. This is in part because conditional heteroskedasticity is discussed in a separate chapter. Random coefficient models also belong under that heading and have not been considered here. Furthermore, this presentation reflects the belief that economic phenomena are more naturally characterized by stochastic than deterministic models, so that deterministic chaos and its applications to economics have only been briefly mentioned in the discussion.

At present the number of applications of nonlinear time series models in economics is still fairly limited. Many techniques discussed here are as yet relatively untested. However, the situation may change rather rapidly, so that in a few years the possibilities of evaluating the empirical success of the present and new techniques will be essentially better than now.

REFERENCES

Akaike, H. (1969). Fitting autoregressions for predictions. Annals of the Institute of Statistical Mathematics 21, 243-247.

Anderson, B.D.O. and J.B. Moore (1979). Optimal filtering. Englewood Cliffs, NJ: Prentice-Hall.

Andrews, D.W.K. (1990). Tests for parameter instability and structural change with unknown change point. Cowles Foundation for Research in Economics, unpublished manuscript.

Auestad, B. and D. Tjøstheim (1990). Identification of nonlinear time series: First order characterization and order determination. Biometrika 77, 669-687.

Auestad, B. and D. Tjøstheim (1991). Functional identification in nonlinear time series. In: G. Roussas, ed. Nonparametric functional estimation and related topics. Amsterdam: Kluwer Academic Publishers, 493-507.

Barnett, W.A., J. Powell and G.E. Tauchen, eds. (1991). Nonparametric and semiparametric methods in econometrics and statistics. Proceedings of the 5th. International Symposium in Economic Theory and Econometrics. Cambridge: Cambridge University Press.

Bates, D.M. and D.G. Watts (1988). Nonlinear regression analysis & its applications. New York: Wiley.

Box, G.E.P. and G.M. Jenkins (1970). Time series analysis, forecasting and control. San Francisco: Holden-Day.

Breiman, L. and J.H. Friedman (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). Journal of the American Statistical Association 80, 580-619.

Breiman, L., J.H. Friedman, R. Olshen and C.J. Stone (1984). Classification and regression trees. Belmont, CA: Wadsworth.

Brock, W.A., Dechert, W.D. and A.J. Scheinkman (1987). A test for independence based on the correlation dimension. Working paper, University of Wisconsin-Madison, Social Systems Research Institute.

Brock, W.A. and S.M. Potter (1992). Nonlinear time series and macroeconometrics. Unpublished manuscript.

Brockett, P.L., M.J. Hinich and D. Patterson (1988). Bispectral-based tests for the detection of Gaussianity and linearity in time series. Journal of the American Statistical Association 83, 657-664.

Brown, R.L., J. Durbin and J.M. Evans (1975). Techniques for testing the constancy of regression coefficients over time. Journal of the Royal Statistical Society B, 37, 149-192 (with Discussion).

Chan, K.S. (1990). Testing for threshold autoregression. Annals of Statistics 18, 1886-1894.

Chan, K.S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. Journal of the Royal Statistical Society B, 53, 691-696.

Chan, K.S. and H. Tong (1986). On estimating thresholds in autoregressive models. Journal of Time Series Analysis 7, 179-190.

Chan, K.S. and H. Tong (1990). On likelihood ratio tests for threshold autoregression. Journal of the Royal Statistical Society B, 52, 469-476.

Chen, P. and R.H. Day, eds. (1992). Non-linear dynamics and evolutionary economics. Cambridge, MA: MIT Press (forthcoming).

Chen, R. and R. Tsay (1990). Nonlinear time series analysis: A functional coefficient autoregressive model approach. Preprint. Department of Statistics, Texas A & M University.

Chen, R. and R. Tsay (1991). Nonlinear additive ARX models. Preprint. Graduate School of Business, University of Chicago.

Cheng, B. and H. Tong (1992). On consistent nonparametric order determination and chaos. Journal of the Royal Statistical Society, B (forthcoming).

Chow, G.C. (1960). Testing for equality between sets of coefficients in two linear regressions. Econometrica 28, 591-605.

Davidson, R. and J.G. MacKinnon (1985). Heteroskedasticity-robust tests in regressions directions. Annales de l'INSEE 59/60, 183-218.

Davies, R.B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 247-254.

De Gooijer, J.G. and K. Kumar (1991). Some recent developments in non-linear time series modelling, testing and forecasting. Unpublished paper.

Desai, M. (1984). Econometric models of the share of wages in national income, U.K., 1855-1965. In: R.M. Goodwin, M. Kruger and A. Vercelli, eds. Nonlinear models of fluctuating growth. Lecture Notes in Economics and Mathematical Systems No. 228. New York: Springer Verlag.

Diebolt, J. (1990). Testing the functions defining a nonlinear autoregressive time series. Stochastic Processes and their Applications 36, 85-106.

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

Engle, R.F. and V. Ng (1991). Measuring and testing the impact of news on volatility. University of California, San Diego, Department of Economics, Discussion Paper No. 91-15.

Ertel, J.E. and E.B. Fowlkes (1976). Some algorithms for linear spline and piecewise multiple linear regression. Journal of the American Statistical Association 71, 640-648.

Franke, J. and M. Wendel (1990). A bootstrap approach for nonlinear autoregressions. Some preliminary results. Preprint, to appear in Proceedings of the International Conference on Bootstrapping and Related Techniques, Trier, June 1990.

Friedman, J.H. (1991). Multivariate adaptive regression splines (with discussion). Annals of Statistics 19, 1-141.

Friedman, J.H. and W. Stuetzle (1981). Projection pursuit regression. Journal of the American Statistical Association 76, 817-823.

Gallant, A.R. (1981). On the bias in flexible functional forms and an essentially unbiased form: The Fourier Flexible Form. Journal of Econometrics 15, 211-245.

Gallant, A.R. (1987). Nonlinear statistical models. New York: Wiley.

Gallant, A.R. and G. Tauchen (1990). A nonparametric approach to nonlinear time series analysis: Estimation and simulation. Preprint, to appear in IMA volumes on Mathematics and its Applications, Springer Verlag.

Gallant, A.R., P.E. Rossi and G. Tauchen (1990). Stock prices and volume. Graduate School of Business, University of Chicago, working paper.

Granger, C.W.J. and J.J. Hallman (1991a). Nonliner transformations of integrated time series. Journal of Time Series Analysis 12, 207-224.

Granger, C.W.J. and J.J. Hallman (1991b). Long-memory processes with attractors. Oxford Bulletin of Economics and Statistics 53, 11-26.

Granger, C.W.J. and J.L. Lin (1991). Nonlinear correlation coefficients and identification of nonlinear time series models. University of California, San Diego, Department of Economics, Discussion Paper.

Granger, C.W.J. and P. Newbold (1986). Forecasting economic time series, 2nd edition. Orlando, FL: Academic Press.

Granger, C.W.J. and T. Teräsvirta (1991). Experiments in modeling nonlinear relationships between time series. Proceedings of the NATO/Santa Fe Conference on Non-

linear Modeling and Forecasting (forthcoming).

Granger, C.W.J. and T. Teräsvirta (1992). Modelling nonlinear dynamic relationships. Oxford: Oxford University Press (forthcoming).

Granger, C.W.J., T. Teräsvirta and H.M. Anderson (1992). Modelling non-linearity over the business cycle. In: J.H. Stock and M.W. Watson, eds. New research on business cycles, indicators and forecasting. Chicago: Chicago University Press (forthcoming).

Haggan, V., S.M. Heravi and M.B. Priestley (1984). A study of the application of state-dependent models in nonlinear time series analysis. Time Series Analysis 5, 69-102.

Haggan, V. and T. Ozaki (1981). Modelling non-linear random vibrations using an amplitude-dependent autoregressive time series model. Biometrika 68, 189-196.

Hallman, J.J. (1990). Nonlinear integrated series, cointegration and application. PhD Thesis. University of California, San Diego, Department of Economics.

Hansen, B.E. (1990). Lagrange multiplier tests for parameter instability in non-linear models. Paper presented at the Sixth World Congress of the Econometric Society, Barcelona.

Härdle, W. (1990). Applied nonparametric regression. Oxford: Oxford University Press.

Harvey, A.C. (1990). Econometric analysis of time series, 2nd edition. Cambridge, MA: MIT Press.

Hastie, T.J. and R.J. Tibshirani (1990). Generalized additive models. London: Chapman and Hall.

Heckman, N. (1986). Spline smoothing in a partly linear model. Journal of the Statistical Society, B, 48, 244-248.

Higgins, M. and A.K. Bera (1989). A joint test for ARCH and bilinearity in the regression model. Econometric Reviews 7, 171-181.

Hinich, M.J. (1982). Testing for gaussianity and linearity of a stationary time series. Journal of Time Series Analysis 3, 169-176.

Joe, H. (1989). Estimation of entropy and other functionals of a multivariate density. Annals of the Institute of Statistical Mathematics 41, 683-697.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl and T.-C. Lee (1985). The theory and practice of econometrics, 2nd edition. New York: Wiley.

Keenan, D.M. (1985). A Tukey non-additivity type test for time series nonlinearity. Biometrika 72, 39-44.

Krämer, W., W. Ploberger and R. Alt (1988). Testing for structural change in dynamic

models. Econometrica 56, 1335-1370.

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

Lasota, A. and M.C. Mackey (1987). Stochastic perturbation of dynamical systems: The weak convergence of measures. Journal of Mathematical Analysis and Applications 138, 232-248.

Lee, T.-H., White, H. and C.W.J. Granger (1992). Testing for neglected nonlinearity in time series models. A comparison of neural network methods and alternative tests. Journal of Econometrics (forthcoming).

Lewis, P.A.W. and J.G. Stevens (1991a). Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). Journal of the American Statistical Association 86, 864-877.

Lewis, P.A.W. and J.G. Stevens (1991b). Semi-multivariate nonlinear modeling of time series using multivariate adaptive regression splines (MARS). Preprint, Naval Postgraduate School.

Liu, T., C.W.J. Granger and W. Heller (1991). Using the correlation exponent to decide if an economic series is chaotic. University of California, San Diego, Department of Economics, Discussion Paper No. 91-21.

Luukkonen, R. (1990). On linearity testing and model estimation in non-linear time series analysis. Helsinki: Finnish Statistical Society.

Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988a). Testing linearity in univariate time series. Scandinavian Journal of Statistics 15, 161-175.

Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988b). Testing linearity against smooth transition autoregression. Biometrika 75, 491-499.

Maddala, G.S. (1977). Econometrics. New York: McGraw-Hill.

Maddala, G.S. (1986). Disequilibrium, self-selection and switching models. In: Z. Griliches and M.D. Intriligator, eds. Handbook of econometrics. Vol. 3. Amsterdam: North-Holland, 1634-1688.

Marron, S. (1989). Automatic smoothing parameter selection: A survey. In: A. Ullah, ed. Semiparametric and nonparametric econometrics. Heidelberg: Physica-Verlag, 65-86.

Müller, H.G. (1990). Smooth optimum kernel estimators near endpoints. Preprint, University of California, Davis.

Nicholls, D.F. and A.R. Pagan (1985). Varying coefficient regression. In: E.J. Hannan, P.R. Krishnaiah and M.M. Rao, eds. Handbook of statistics, Vol. 5. Amsterdam: Elsevier, 413-449.

Time Series Analysis 4, 185-208.

Robinson, P.M. (1988). Root-N-consistent semiparametric regression. Econometrica 56, 931-954.

Robinson, P.M. (1991). Consistent nonparametric entropy-based testing. Review of Economic Studies (forthcoming).

Saikkonen, P. and R. Luukkonen (1988). Lagrange multiplier tests for testing non-linearities in time series models. Scandinavian Journal of Statistics 15, 55-68.

Scheinkman, J.A. (1990). Nonlinearities in economic dynamics. Economic Journal 100, Supplement, 33-48.

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

Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. Annals of Statistics 12, 898-916.

Stensholt, B.K. and D. Tjøstheim (1987). Multiple bilinear time series models. Journal of Time Series Analysis 8, 221-233.

Stinchcombe, M. and H. White (1989). Universal approximations using feedforward networks with non-sigmoid hidden layer activation functions. Proceedings of the International Joint Conference on Neural Networks, Washington, D.C. San Diego: SOS Printing, I: 613-618.

Subba Rao, T. and M.M. Gabr (1980). A test for linearity of stationary time series. Journal of Time Series Analysis 1, 145-158.

Subba Rao, T. and M.M. Gabr (1984). An introduction to bispectral analysis and bilinear time series models. Lecture Notes in Statistics 24, New York: Springer.

Sughihara, G. and R.M. May (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344, 734-741.

Teräsvirta, T. (1990a). Specification, estimation, and evaluation of smooth transition autoregressive models. University of California, San Diego, Department of Economics, Discussion Paper No. 90-39.

Teräsvirta, T. (1990b). Power properties of linearity tests for time series. University of California, San Diego, Department of Economics, Discussion Paper No. 90-15.

Teräsvirta, T. (1990c). Generalizing threshold autoregressive models. University of California, San Diego, Department of Economics, Discussion Paper No. 90-44.

Teräsvirta, T. and H.M. Anderson (1991). Modelling nonlinearities in business cycles

Yakowitz, S. (1987). Nearest-neighbour methods for time series analysis. Journal of Time Series Analysis 8, 235-247.

1990:1	Holm, S.	Abstract bootstrap confidence intervals in linear models.
1990:2	Holm, S. & Dahlbom, U	On tests of equivalence
1991:1	Olofsson, Jonny	On some prediction methods for categorical data
1991:2	Jonsson, Robert	On the problem of optimal inference in the simple error component model for panel data
1991:3	Holm, S. & Svensson, E.	Statistical rank methods for ordinal categorical data
1992:1	Teräsvirta, T Tjøstheim, D Granger, C W J	Aspects of Modelling Nonlinear Time Series