Volume Title: Business Cycles, Indicators and Forecasting

Volume Author/Editor: James H. Stock and Mark W. Watson, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-77488-0

Volume URL: http://www.nber.org/books/stoc93-1

Conference Date: May 3-4, 1991

Publication Date: January 1993

Chapter Title: Modeling Nonlinearity over the Business Cycle

Chapter Author: Clive W. Granger, Timo Terasvirta, Heather M. Anderson

Chapter URL: http://www.nber.org/chapters/c7196

Chapter pages in book: (p. 311 - 326)

8 Modeling Nonlinearity over the Business Cycle

Clive W. J. Granger, Timo Teräsvirta, and
Heather M. Anderson

There seems to be wide acceptance that the relations between economic time series are often nonlinear. This is certainly true for relations suggested by economic theory, and econometricians can propose estimates of the parameters of any explicitly nonlinear model. However, exploratory or specification search forms of modeling are used when a theory is not specific about the dynamics of a relation apply to linear (or log-linear) models, such as the Box-Jenkins transfer functions and vector autoregressive models. There is no generally accepted class of nonlinear models that can be applied to explore relations. One difficulty is that the number of alternative nonlinear models is enormous and that there has not been sufficient experience accumulated to decide which of these models are most appropriate in economics. In the past decade or so, there has been a greater deal of activity among statisticians, time-series analysts, and econometricians suggesting possible nonlinear models and techniques for their analysis. Tong (1990) surveys many of the univariate models, and Granger and Teräsvirta (in press) consider multivariate models but concentrate on single-equation nonlinear relations. The possible models include parametric forms (such as nonlinear autoregressive, bilinear, and doubly stochastic models), state space and flexible Fourier forms, many types of nonparametric models (including projection pursuit), and mixtures of these called semiparametric. These models are designed to be fairly general and to be flexible so that they can approximate a wide variety of actual nonlinearities. A problem that arises from this flexibility is that the models are in-

Clive W. J. Granger is professor of economics at the University of California, San Diego. Timo Teräsvirta is a research fellow at the Bank of Norway and visiting professor of economics at the University of California, San Diego. Heather M. Anderson is assistant professor of economics at the University of Texas at Austin.

This research has been funded by National Science Foundation grant SES-89-02950 and by the Yrjö Jahnsson Foundation.

clined to overfit in sample. We thus believe that it is important first to test linearity against the nonlinearity in which we are interested and then only if linearity is rejected to build a nonlinear model. Further, one way of judging the quality of this model should be its out-of-sample performance compared to other models, although there can be difficulties with this strategy, as will be seen later.

There exist many tests of linearity, and many of these are discussed in Lee, White, and Granger (1993), where a variety of simulations are presented, and in Teräsvirta (1990a), where theoretical power properties are discussed. A few tests are found to have good power under a variety of situations, including a new test based on neural network models. Let $\phi(X)$ be a so-called squashing function, being smooth, bounded, and monotonic nondecreasing, such as a probability density function or a logistic function, so that in this last case

$$Y_t = c + \beta' \underline{X}_t + \sum_{j=1}^{p} \alpha_j \phi(\gamma_j' \underline{X}_t) + \text{white noise}, \tag{1}$$

where $\underline{X}_t$ is the vector of explanatory variables including lagged Y's and present and lagged values of other variables. It is, of course, rather complicated to estimate the γ parameters, so a simple procedure has been suggested by White (1989) in which values of $\underline{\gamma}$ are chosen at random from some appropriate region and a rather large value of p is used, say 10 or 20. As the ϕ terms are now directly observable, $\underline{\beta}$ and the α's can be estimated by a standard regression procedure and the significance of the α's tested directly. If any are found to be significant, linearity can be rejected. In Lee, White, and Granger (1993), this test was found to have good power in most cases, but not for data generated by a bilinear model.

In Teräsvirta, Lin, and Granger (1993), it is proposed that a Lagrange multiplier (LM)–type test be used. As expected, simulations showed that it has comparable power to the neural network test and often better power. The test, here called the polynomial test, involves just adding squares and cross-products of the components of $\underline{X}_t$ (quadratic terms) and possibly also cubic terms (such as X_i^3, $X_i^2X_j$, and $X_iX_jX_k$). These are added to a linear model, with constant, to form an artificial regression and an F-test used.

If such a test suggests that linearity can be rejected, the question naturally arises of what model to fit to the data. The polynomial test immediately suggests a model, involving the quadratic and cubic terms, but this form is not easily interpreted and may be explosive. It also suggests (1) because the test is an LM-type test against this alternative. However, difficulties in estimating (1) discourage use of that model.

The class of models that we propose is known as smooth transition regressions (STRs), which take the form

$$Y_t = c_1 + \underline{\beta}_1 \underline{X}_t + \phi(Z_t)(c_2 + \underline{\beta}_2' \underline{X}_t) + e_t, \tag{2}$$

where $0 \leq \phi(Z) \leq 1$, and $\phi(Z)$ is zero, or small, for some values of Z and is near to one for other Z values. In (2), Z_t is the "indicator variable" and may be a linear combination of the components of $\underline{X}_t$ or just a single lagged component of $\underline{X}_t$, plus a constant, for example, $Z_t = m + \gamma Y_{t-d}$. The model is seen to be a smooth transition between the linear model

$$Y_t = c_1 + \underline{\beta}_1' \underline{X}_t + e_t$$

and the alternative linear model

$$Y_t = (c_1 + c_2) + (\underline{\beta}_1 + \underline{\beta}_2)' \underline{X}_t + e_t.$$

These two linear models can have quite different properties—one could be I(1) and the other I(0), for example—and can perhaps be interpreted as two regimes. The models thus represent a smooth regime-switching situation. There are plenty of examples in economic theory of regime switching, such as full-employment or full-capacity models being different from the nonfull cases. Similarly, there is evidence of stock market price changes being forecastable for certain indicator variable values, such as low volatility, but not otherwise. The following section considers the specification and testing of these models.

8.1 Specification of Smooth Transition Regression Models

Consider the following smooth transition regression (STR) model:

$$(3) \qquad y_t = \beta' x_t + (\theta' x_t) F(\alpha' z_t) + u_t,$$

where $u_t \sim$ i.i.d$(0, \sigma^2)$, $E(z_t u_t) = 0$, $\beta = (\beta_0, \beta_1, \ldots, \beta_m)'$, $\theta = (\theta_0, \theta_1, \ldots, \theta_m)'$, $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{m+h})'$, $\Sigma_{j=10}^{m+h} \alpha_j = 1$, say, $x_t = (1, y_{t-1}, \ldots, y_{t-p}, x_{1t}, \ldots, x_{kt})'$, and $z_t = (x_t', v_t')'$ with $v_t = (v_{t1}, \ldots, v_{th})'$. Two different types of F will be considered. First,

$$(4) \qquad F(\alpha' z_t) = \{1 + \exp[-\gamma(\alpha' z_t)]\}^{-1}, \quad \gamma > 0,$$

which makes (3) a logistic STR (LSTR) model. Maddala (1977, 396) suggested this formulation. If $\gamma = 0$, (3) is a linear model. Second, if

$$(5) \qquad F(\alpha' z_t) = 1 - \exp[-\gamma(\alpha' z_t)^2], \quad \gamma > 0,$$

we have an exponential STR (ESTR) model. If $\gamma \to \infty$ in (4), (3) becomes a switching regression model with $\alpha' z_t$ as the linear combination of transition variables. If $\gamma \to \infty$ in (5), (3) becomes a linear model; its parameters switch if $\alpha' z_t = 0$, but that is an event with zero probability. Functions (4) and (5) describe two fundamentally different forms of parameter behavior. If (4) holds, the "stochastic parameter vector" $\beta + \theta F$ in (3) changes monotonically from β to $\beta + \theta$ with $\alpha' z_t$. If (5) is valid, this change is symmetrical about zero: the parameters change from $\beta + \theta$ to β and back again with increasing $\alpha' z_t$. Usually, (4) and (5) are still too general, in particular if the time series available for modeling are relatively short. A way of restricting them is to let $\alpha = (-c, 0, \ldots, 0, 1, 0, \ldots, 0)'$ so that $\alpha' z_t = z_{tj} - c$. Instead of a linear combination of variables determining the transition, a single variable is doing it. In practice, this variable may often be unknown; that is, it is not known to the investigator which $\alpha_j = 1$, and that has to be determined from the data. Furthermore, it is assumed here that, if the model is nonlinear, it is not known if the transition function is (4) or (5).

A successful application of (3) to data requires the following three steps: (i) Specify a linear model to form a basis for further analysis. (ii) Test linearity of (3) ($\gamma = 0$) against nonlinearity ($\gamma > 0$). If linearity is rejected, determine the transition variable from the data. Once this has been done, (iii) the transition function has to be selected, the choice being between (4) and (5). These steps may be carried out as follows:

i. Carry out the complete specification of a linear model. The maximum lag length p for lagged y_t has to be determined from the data as well as regressors $x_{t1}, \ldots, x_{tk}$ if economic theory is not fully explicit about them.
ii. Test linearity of (3) against STR using each element of z_t in turn as the transition variable. If linearity is rejected for more than one transition variable, choose the one for which the p-value of the test is the lowest. The rationale behind this procedure is that the linearity test has the highest power against the correctly specified alternative. If a wrong transition variable is selected, the power of the test suffers from the erroneous choice.
iii. Treat the selected transition variable as given, and choose between (4) and (5) using an auxiliary regression already needed in testing linearity.

It is seen that linearity testing plays a central role in this specification strategy. In this case, not the only, but the most convenient, way of expressing the null hypothesis of linearity is

$$H_0^*: \gamma = 0 \quad \text{against} \quad H_1^*: \gamma > 0. \tag{6}$$

It is clear from (3) that the model is not identified under this null hypothesis. A Lagrange multiplier–type test may be derived for testing (6) using the suggestions in Davies (1977). The test is described in detail in Granger and Teräsvirta (in press, chaps. 6 and 7). It is based on the following artificial regression:

$$y_t = \delta_0 + \delta_1' \tilde{x}_t + \delta_2'(\tilde{x}_t z_{td}) + \delta_3'(\tilde{x}_t z_{td}^2) + \delta_4'(\tilde{x}_t z_{td}^3) + \tilde{u}_t, \tag{7}$$

where $\tilde{u}_t \sim$ i.i.d.$(0, \sigma^2)$, $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-p}, x_{1t}, \ldots, x_{kt})'$ and z_{td} is the transition variable, which is an element of $\tilde{x}_t$. The null hypothesis (6) translates into

$$H_0: \delta_2 = \delta_3 = \delta_4 = 0. \tag{8}$$

Table 8.1 **Alternative Choices of STR Model Type Based on Outcomes of a Sequence of F-Tests within the Artificial Model (7)**

Hypothesis			
$\delta_4 = 0$	$\delta_3 = 0 \mid \delta_4 = 0$	$\delta_2 = 0 \mid \delta_3 = \delta_4 = 0$	Choice
Reject	. . .	. . .	LSTR
Accept	Reject	Accept	ESTR
Accept	Accept	Reject	LSTR
Accept	Reject	Reject	No decision

The alternative is that at least one element in either δ_2, δ_3, or δ_4 is not equal to zero. The fourth-order terms are needed in the case that the true model is an LSTR model in which the most important nonlinearity parameters is the intercept in the nonlinear part, θ_0.

If the true model is an ESTR model, then $\delta_4 = 0$ (see Granger and Teräsvirta, in press, chap. 7). Also, if $\delta_3 = 0$, the model can be only an LSTR model. This suggests the following testing sequence. First, test H_0 in (7), and continue as follows:

i. If (8) is rejected, test H_{04}: $\delta_4 = 0$ against H_{14}: $\delta_4 \neq 0$ in (7).
ii. If H_{04} is accepted, test H_{03}: $\delta_3 = 0 \mid \delta_4 = 0$ against H_{13}: $\delta_3 \neq 0 \mid \delta_4 = 0$.
iii. If H_{03} is accepted, test H_{02}: $\delta_2 = 0 \mid \delta_3 = \delta_4 = 0$ against H_{12}: $\delta_2 \neq 0 \mid \delta_3 = \delta_4 = 0$.

In practice, it is advisable to carry out these three tests automatically independent of the outcome of the previous test. The outcomes help us decide between LSTR and ESTR models (see table 8.1).

The only problem is the case where both H_{03} and H_{02} are rejected after H_{04} is accepted. Then this scheme does not provide a clear-cut answer to the problem of choosing between the two models (see Granger and Teräsvirta, in press chap. 7). Recomputing the test statistic after shifting z_{td} may provide some guidance, as in the univariate case that Teräsvirta (1990b) discussed.

The situation is less complicated if the transition variable is not an element of $\tilde{x}_t$ but one of $v_t = (v_{t1}, \ldots, v_{th})'$, say, v_{td}. Then an artificial regression corresponding to (7) is

$$y_t = \delta_0 + \delta_1'\tilde{x}_t + \delta_2 v_{td} + \delta_3 v_{td}^2 + \delta_4'\tilde{x}_t v_{td} + \delta_5'\tilde{x}_t v_{td}^2 + \tilde{u}_t.$$

The null hypothesis of linearity is

$$(9) \qquad H_0\text{: } \delta_2 = \delta_3 = 0;\ \delta_4 = \delta_5 = 0,$$

the alternative being that (9) is not valid. The test has power against both LSTR and ESTR alternatives. If, after rejecting (9), we test H_{05}: $\delta_5 = 0$ and reject it, the conclusion is that (3) is an ESTR model.

The steps of the specification strategy outlined above leave open the question of the dynamic structure of (3). Specifying that may be best carried out by estimating different specifications and learning from them. This has turned out to be a successful way in univariate STAR (smooth transition autoregression) modeling (for examples, see Teräsvirta 1990b; and Teräsvirta and Anderson 1992). The specification of STAR models follows the same principles as those presented above. The main difference is that $x_t = (1, y_{t-1}, \ldots, y_{t-p})'$ and the transition variable is $z_{td} = y_{t-d}$ where d is not known a priori (for discussion, see Teräsvirta 1990b).

8.2 Modeling the Relations between GNP and the Index of Leading Indicators

This section describes an exploratory modeling exercise between y_t = real GNP and x_t = the Department of Commerce quarterly index of leading indicators. The objective of the exercise is to consider if a nonlinear model provides better forecasts of GNP. The results are indecisive, with some evidence of nonlinearity being found but no clear-cut improvement in forecastability. The series used are those designated GNP82 and DLEAD in the Citibase data bank. The series are quarterly, real, and seasonally adjusted, with 166 observations from the period 1948: I–1989: II. Models were constructed using the full sample, and then twenty terms were held back for use in a forecast comparison, so that the models were reestimated using just the first 146 observations. The leading indicator series was originally recorded monthly but was summed over the adjacent values to obtain a quarterly series. It was decided to use GNP as the variable of interest as it is probably the best available approximation to the variable that the leading indicators were designed to lead. The problem is that this variable is available only quarterly. The alternative would be to use the index of industrial production, which is available monthly, but, with the growth in the importance of the service industries, the industrial sector now provides a poor approximation of GNP.

To help decide what models to fit, the two series, y_t and x_t, and their logs, Ly_t and Lx_t, were tested for unit roots and then for cointegration.

Augmented Dickey-Fuller τ-tests (see Dickey and Fuller 1979) suggest that both y and x are I(1), that Lx is also I(1), but that Ly is less clearly so. (Details are shown in appendix 8A.1.) The first-differences of all these series are all clearly I(0). The fitted univariate linear model for y_t is

$$\text{(10)} \qquad \underset{}{y_t} = \underset{(.08)}{1.28y_{t-1}} - \underset{(.13)}{0.14y_{t-2}} - \underset{(.08)}{0.14y_{t-3}} + \underset{(5.67)}{3.64} + e_t,$$

$$R^2 = .9992, \text{D-W} = 1.97, \tau_0 = -12.22$$

where standard errors are shown in brackets, and τ_k is the augmented Dickey-Fuller τ-statistic applied to the residuals using k lags of the differenced variable. It is seen that the coefficients add to one, and the errors e_t seem to be I(0) (or stationary), suggesting that y_t is I(1), possibly with drift.

To investigate possible cointegration, it will be assumed that all the series are I(1). The Engle-Granger (1987) two-step tests found no evidence of cointegration in levels and weak evidence of cointegration for logs, as the following regressions show:

$$(11)\quad \underset{}{y_t} = \underset{(25.6)}{520.3} + \underset{(.52)}{9.10t} + \underset{(.27)}{4.57x_t}, \quad \text{D-W} = .133,\ \tau_1 = -2.82;$$

$$(12)\quad y_t = \underset{(25.3)}{155.5} + \underset{(.10)}{9.27x_t}, \quad \text{D-W} = .15,\ \tau_4 = -3.35;$$

$$(13)\quad Ly_t = \underset{(.19)}{5.07} + \underset{(.0003)}{.004t} + \underset{(.04)}{.421Lx_t}, \quad \text{D-W} = .11,\ \tau_1 = -4.02;$$

$$(14)\quad Ly_t = \underset{(.05)}{2.59} + \underset{(.009)}{.85Lx_t}, \quad \text{D-W} = .18,\ \tau_4 = -3.61.$$

The critical value for the τ-statistics is -3.44, so the null hypothesis that the residuals are I(1) is rejected only in (13) and (14). It should be noted that the standard errors given here cannot be used to evaluate the significance of parameters in the model because of the very low Durbin-Watson statistics. However, these results are confused by different results obtained by using the Johansen (1988) maximum likelihood procedure, which find cointegration for both levels and logarithms of the series. (Again, details are shown in appendix table 8A.2.) The Johansen procedure is known to be more powerful than the two-step procedure (as shown by Gonzalo [1991]), but the results do not suggest that it is necessarily a good modeling strategy to impose cointegration on further models, so a flexible specification was adopted.

For later purposes, it is relevant to ask if various nonlinear transforms of our series seem to be I(1) or not. Using both simulations and simple theory, Granger and Hallman (1988) show that polynomials in I(1) variables also contain linear roots if a linear model is constructed. Table 8.2 shows Dickey-Fuller τ-test statistics for several of these polynomials. A critical value of -3.44 or less allows one to reject the null hypothesis of a unit root with 95 percent confidence. It is seen that, using x, y, in no case is the null rejected, but it is rejected twice using Lx, Ly.

The modeling experiment used the following steps:

i. Construct a few alternative models for y_t with and without the constraints suggested by possible cointegration, using the full sample.
ii. Use Lagrange multiplier tests to test for missing variables that are second- and third-order polynomials of lagged dependent (GNP) and explanatory (leading index) variables.
iii. Reestimate the model using the shorter sample, and compare the one-step forecasting ability of the alternative models over the final twenty terms of the sample.

Table 8.2 **Augmented Dickey-Fuller (D-F) Tests for Quadratic and Cubic Functions o x and y**

Variable	y^2	xy	x^2	y^3	y^2x	yx^2	x^3
D-F stat.	.59	−.17	−.77	1.93	.79	.64	.10
Variable	$(Ly)^2$	Lx,Ly	$(Lx)^2$	$(Ly)^3$	$(Ly)^2Lx$	$Ly(Lx)^2$	$(Lx)^3$
D-F stat.	−3.01	−3.17	−3.26	−3.71	−4.33	−3.19	−3.19

These steps were repeated for the logarithms of the variables, Ly_t and Lx_t. Once evidence for nonlinearity has been found, the best procedure is to try to build specific models, such as nonlinear autoregressive or neural network models, but this has been attempted just with an STR model, as reported in the next section.

Two of the preliminary models considered for y_t are reported (together with specifications tests and the Lagrange multiplier test results, for possible augmenting nonlinear variables). Denoting these models as M1 and M2, M1 is given by,

$$\begin{aligned}\Delta y_t = \underset{(.08)}{.06\Delta y_{t-1}} + \underset{(2.04)^*}{12.55} + \underset{(.35)^*}{2.45\Delta x_{t-1}},\end{aligned} \tag{15}$$

$N = 164$ (1948:III–1989:II),

$R^2 = .33$, D-W $= 2.14$, $\tau_1 = -7.78^*$, SE $= 21.17$,

Serial Correlation: $F(6,155) = 1.26$, ARCH: $F(6, 149) = .52$, Normality: $\chi^2 = 24.51^*$,

$\text{LM}(\Delta y)_{-1}^2\,(\Delta x)_{-1}$: $F(1, 160) = 5.17^*$,

and M2 is given by

$$\Delta y_t = \underset{(1.75)^*}{14.90} - \underset{(.02)^*}{.06(\text{ECT})_{-1}} + \underset{(.31)^*}{(2.05)\Delta x_{t-1}}, \tag{16}$$

$N = 159$ (1949:IV–1989:II),

$R^2 = .38$, D-W $= 2.08$, $\tau_0 = -13.41^*$, SE $= 20.38$,

Serial Correlation: $F(6, 150) = .72$, ARCH: $F(6, 144) = .67$, Normality: $\chi^2 = 18.76^*$,

$\text{LM}(\Delta y_{-1}^2\,\Delta x_{-1}) = F(1, 155) = 5.31^*$.

Here ECT is the error-correction term from (12), ECT $= y_t - 155.5 - 9.27x_t$.

We provide a variety of statistics to help evaluate the models, and these include R^2, Durbin-Watson (D-W), the standard error (standard deviation of

Table 8.3 **In Sample and Postsample Standard Errors for Models M1 and M2 and Their Nonlinear Extension**

	Standard Error		
Model	Long Sample	Short Sample	Postsample
M1	21.17	20.83	24.47
M1 augmented		20.75	23.11
M2	20.38	20.33	21.13
M2 augmented		20.10	20.50

residuals), and τ_k, the augmented Dickey-Fuller test statistic based on the residuals using k lags. Standard specification test statistics for serial correlation, ARCH, and normality are also given. The Lagrange multiplier test results, with the apparently successful augmenting nonlinear variables, are shown after the specification statistics.

A linear trend was added to M1 but was not found to be significant. In both models, the LM tests suggested that some nonlinearity in mean may be present. These models were augmented by adding a term $(\Delta y_{t-1})^2(\Delta x_{t-1})$. Table 8.3 shows the standard errors in sample for both the long and the shortened sample plus the postsample forecast standard errors.

Using the test outlined in Granger and Newbold (1986, 278–79) to compare the sum of squared forecast errors, the M1 augmented model is significantly better than M1. However, the error correction model M2, which is superior both in and out of sample, does not produce a clearly superior forecast when augmented by a nonlinear term.

A more extensive experiment was conducted using the log series and three models, denoted L1, L2 and L3, are reported here. L1 is given by

$$\begin{aligned} Ly_t = {} & \underset{(.02)^*}{.85Ly_{t-1}} + \underset{(.13)^*}{.56} + \underset{(.0001)^*}{.0003t} + \underset{(.01)}{.11Lx_{t-1}}, \end{aligned} \tag{17}$$

$N = 165$ (1948:II–1989:II),

$R^2 = .9994$, D-W $= 1.519$, $\tau_2 = -7.94^*$, SE $= .0093$,

Serial Correlation: $F(6, 155) = 2.78^*$, ARCH: $F(6, 149) = 2.69^*$, Normality: $\chi^2 = 1.17$,

$\text{LM}(LxLy)_{-1}$: $F(1, 160) = 4.55^*$, LM $(Ly)^3_{-1}$: $F(1, 160) = 3.91^*$,

LM $(Ly)^2_{-1}(Lx)_{-1}$: $F(1, 160) = 4.42^*$, LM $(Ly)_{-1}\,(Lx)^2_{-1}$:

$F(1, 160) = 4.68^*$, $\text{LM}(Lx)^3_{-1}$: $F(1, 160) = 4.77^*$.

L2 is given by

$$(18)\quad \underset{}{\Delta Ly_t} = \underset{(.07)}{.11\Delta Ly_{t-1}} + \underset{(.0009)^*}{.005} + \underset{(.03)^*}{.23\Delta Lx_{t-1}},$$

$N = 164$ (1948:III–1989:II),

$R^2 = .34$, D-W $= 2.19$, $\tau_0 = -13.24^*$, SE $= .0090$,

Serial Correlation: $F(6, 155) = 1.40$, ARCH: $F(6, 149) = 1.69$, Normality: $\chi^2 = 4.58^*$,

$\text{LM}[(\Delta Ly)^2_{-1}, (\Delta Ly\Delta Lx)_{-1}, (\Delta Lx)^2_{-1}]$: $F(3, 158) = 4.30^*$,

$\text{LM}[(\Delta Ly)^3_{-1}, (\Delta Ly)^2_{-1}(\Delta Lx)_{-1}, (\Delta Ly)_{-1}(\Delta Lx)^2_{-1}, (\Delta Lx)^3_{-1}$:

$F(4, 157) = 3.08^*$,

and, finally, L3,

$$(19)\quad \Delta Ly_t = \underset{(.0007)^*}{.007} - \underset{(.016)^*}{.08(\text{ECT})_{-1}} + \underset{(.03)^*}{.18\Delta Lx_{t-1}},$$

$N = 159$ (1949:IV–1989:II),

$R^2 = .42$, D-W $= 1.99$, $\tau_6 = -4.22^*$, SE $= .0084$,

Serial Correlation: $F(6, 150) = .73$, ARCH: $F(6, 144) = .24$, Normality: $\chi^2 = 2.49^*$,

$$\text{LM}\left[(\Delta Ly)^2_{-1}, (\Delta Ly_{-1})(\Delta Lx)_{-1}, (\Delta Lx)^2_{-1}\right]: F(3, 153) = 3.79^*.$$

In each of the models, the LM tests found evidence of nonlinearity. Table 8.4 shows the in-sample and postsample values of the standard errors of the various models.

It is seen that the augmentations of L1 all involve particular single quadratic and cubic terms and that all the augmented models forecast better than the original linear models, and the improvement in forecasting ability is found to be significant. The best linear model, both in and out of sample, is L3, which is an error-correction form, and the augmentation with quadratic lagged terms gives a slight improvement. Thus, the initial conclusion is that there does appear to be some evidence of nonlinearity in mean and that the nonlinear models can lead to an improvement in forecasts.

The augmented models are considered just for the purpose of testing for nonlinearity and are not proposed as serious, interpretable, nonlinear models. Their postsample performance is presented just for general interest.

One final experiment included the current value of the index of leading indicators in the model for GNP, in which case the LM tests found no evidence of missing nonlinearity in mean using lagged quadratic or cubic terms. Although this result has no direct forecasting implications for GNP, it suggests that a nonlinear model for the leading index is worth exploring. A modeling

Table 8.4 In-Sample and Postsample Standard Errors for Models L1, L2, and L3

Model	Additional Nonlinear Terms	In Sample	Postsample
L1		.0096	.0076
L1A$_1$	$(Lx_{-1}\ Ly_{-1})$	.0096	.0057
L1A$_2$	$(Ly_{-1})^3$	.0096	.0061
L1A$_3$	$(Ly_{-1})^2(Lx_{-1})$	.0096	.0058
L1A$_4$	$(Ly_{-1})(Lx_{-1})^2$	.0096	.0056
L1A$_5$	$(Lx_{-1})^3$	.0096	.0055
L2		.0094	.0054
L2A$_1$	$(\Delta Ly_{-1})^2$, $(\Delta Ly_{-1})(\Delta Lx_{-1})$, $(\Delta Lx_{-1})^2$	.0091	.0059
L2A$_2$	$(\Delta Ly_{-1})^3$, $(\Delta Ly_{-1})^2(\Delta x_{-1})$, $(\Delta Ly_{-1})(\Delta Lx_{-1})^2$, $(\Delta Lx_{-1})^3$	.0091	.0061
L3		.0087	.0055
L3A	$(\Delta Ly_{-1})^2$, $(\Delta Ly_{-1})(\Delta Lx_{-1})$, $(\Delta Lx_{-1})^2$	.0085	.0054

exercise, not reported here in detail, did find that a univariate LSTAR model of the index fitted better in sample than a linear model but did not forecast better. It should be noted that we could not find a suitable univariate STAR model for GNP.

8.3 Application of STR Models

The tests of linearity suggest that nonlinear models are appropriate, and this section presents and evaluates a pair of models for the changes in log GNP. A critical decision in the specification of these models is the choice of the switching variable. From the justification of these models presented earlier, this variable should itself be slowly changing, and it should not contain a dominant deterministic trend as the variable is inserted into a bounded function. The variable selected was the linear detrended log of the index of leading indicators, denoted by

$$\mathrm{LI}_t = \log\ (\text{index leading indicator}) - 4.73 - .008t.$$

A linear error-correction model which relaxes Ly_t and Lx_t was estimated to be

$$\begin{aligned}(20)\quad \Delta Ly_t = {} & \underset{(.002)}{.006} - \underset{(.02)}{.09}\mathrm{ECT}_{-1} - \underset{(.08)}{.08}\,\Delta Ly_{t-1} \\ & + \underset{(.08)}{.10}\Delta Ly_{t-2} - \underset{(.08)}{.11}\Delta Ly_{t-3} - \underset{(.08)}{.03}\Delta Ly_{t-4} \\ & + \underset{(.07)}{.02}\Delta Ly_{t-5} + \underset{(.07)}{.11}\Delta Ly_{t-6} + \underset{(.07)}{.06}\Delta Ly_{t-7}\end{aligned}$$

$$+ \underset{(.04)}{0.22}\Delta Lx_{t-1},$$

$$\mathrm{SE} = 0.0085,$$

where $\mathrm{ECT}_t = Ly_t - 2.50 - .96\,Lx_t$. The model is clearly overparameterized, but the standard reduction procedure has not been used.

Testing this error correction model against STR alternatives that use lags of the log of the leading indicator index as indicator variables found strong evidence of nonlinearity when $z_t = \mathrm{LI}_{t-2}$. (The p value of this test was $p = .001$.) The corresponding STR model was estimated to be
.001.) The corresponding STR model was estimated to be

$$\Delta Ly_t = \underset{(.002)}{.011} - \underset{(.02)}{.16}\mathrm{ECT}_{-1} - \underset{(.09)}{.27}\Delta Ly_{t-1} \tag{21}$$

$$- \underset{(.08)}{.17}\Delta Ly_{t-3} - \underset{(.08)}{.13}\Delta Ly_{t-4} + \underset{(.07)}{.13}\Delta Ly_{t-6} + \underset{(.06)}{.010}\Delta Ly_{t-7} + \underset{(.04)}{.19}\Delta Lx_{t-1}$$

$$+ \left[- \underset{(.003)}{.015} + \underset{(.06)}{.14}\mathrm{ECT}_{-1} + \underset{(.19)}{.54}\Delta Ly_{t-1} + \underset{(.20)}{.44}\Delta \mathrm{Ly}_{t-3} + \underset{(.17)}{.32}\Delta \mathrm{Ly}_{t-4}\right]$$

$$\left\{1 + \exp[-221(\mathrm{LI}_{-2} - \underset{(.005)}{.036})]\right\}^{-1},$$

$$\mathrm{SE} = 0.007,$$

There are seen to be quite substantial differences in the two regimes, which approximately correspond to the troughs and peaks of the business cycle. The models were constructed using just the in-sample data. The forecasting performance of the two models (20) and (21) over the following twenty-six quarters (i.e., 1984:III–1990:IV) is presented in table 8.5.

It is noteworthy that both these models produce forecasts that are too high, giving negative forecast errors on every occasion except for three. Neither model performed satisfactorily in forecasting the most recent downturn, as shown in figure 8.1. Figure 8.2 shows the movement of the detrended leading indicator LI over time, which is the switching variable used in the model. Using a lag of two, it is seen that the switching variable has changed little

Table 8.5 **In-Sample and Postsample Standard Errors for Models (20) and (21)**

	Root Mean Squared Errors	
	In Sample	Postsample
Linear (20)	.0085	.0072
STR (21)	.0077	.0109

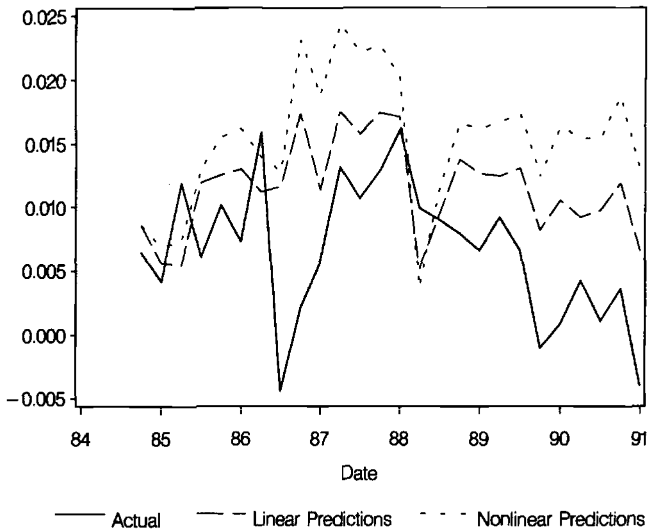

Fig. 8.1 Postsample forecasting performance of the model

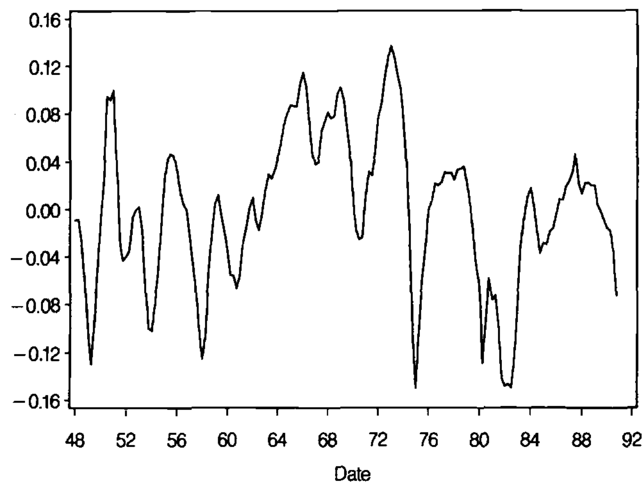

Fig. 8.2 Detrended leading indicator

from its recent high at the end of the postsample period and so was not helpful in forecasting the downturn.

8.4 Conclusions

A number of linear models have been compared to nonlinear ones. It has often been found that the nonlinear models appear to be superior in sample but not out of sample. When using nonlinear models, there is always the possibility of overfitting due to data mining, although this is less likely to occur when tests for linearity have first found evidence against it before model estimation. The postsample evaluation of models is thus particularly important for distinguishing real versus spurious nonlinearity. However, a practical question arises that the postsample period will be limited in extent, and, in this period, both regimes in an STR model may not occur sufficiently often for the inherent advantage of a true nonlinear model to become apparent. For example, if one of the regimes occurs when z_t takes a large value, as in the STR example given above, but this large value occurs less often in the postsample evaluation period, again as in the example, the advantage of the nonlinearity will not be apparent.

Appendix

Table 8A.1 shows the augmented Dickey-Fuller τ_t-test statistics, together with the number of lags used. The 5 percent critical value is -3.44.

The lag k was chosen in each case so that the residual in the Dickey-Fuller equation

$$\Delta y_t = a + by_{t-1} + \gamma t + \sum_{j=1}^{k} \beta_j \Delta y_{t-j} + \text{residual}$$

appears to be white noise. The Dickey-Fuller test statistic is the t-value for the estimate of b.

The Johansen maximum likelihood procedure uses as an alternative test statistic the maximum eigenvalue (λ) and the value of the trace (τ) of an estimated matrix. Table 8A.2 shows the values of these statistics for levels and logs of the variables. The notation r represents the number of cointegrating vectors, and, with only two variables (x and y), r can take only the values zero or one.

The 5 percent critical values for λ_1 and λ_0 are 4.08 and 14.60 and those for τ_1 and τ_0 8.08 and 17.84, respectively (Johansen and Juselius 1990). The evidence suggests that there exists a single cointegrating vector in both levels and logs.

Table 8A.1 **Augmented Dickey-Fuller Test Statistics**

Variable	No. of Lags Used	D-F Test Statistic	Variable	No. of Lags Used	D-F Test Statistic
y	2	-1.69	Δy	0	-8.98
x	4	-2.13	Δx	3	-7.56
Ly	2	-3.71	ΔLy	2	-6.67
Lx	4	-3.16	ΔLx	3	-8.14

Table 8A.2 **Maximum Eigenvalue and Trace Statistics**

	Series in Levels			Series in Logs	
Null	λ	τ	Null	λ	τ
$r \leq 1$	2.03	2.03	$r \leq 1$	1.37	1.37
$r = 0$	35.70	37.73	$r = 0$	40.09	41.46

References

Davies, R. B. 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64:247–54.

Dickey, D., and F. W. Fuller. 1979. Distribution of the estimators for autoregressive time-series with a unit root. *Journal of the American Statistic Association* 74:427–31.

Engle, R. F., and C. W. J. Granger. 1987. Cointegration and error-correction representation, estimation and testing. *Econometrica* 55:251–76.

Gonzalo, J. 1991. Estimation of long-run equilibrium relationships and common long memory components in cointegrated systems. Ph.D. diss., University of California, San Diego, Economics Department.

Granger, C. W. J., and J. Hallman. 1988. The algebra of I(1). Finance and Economics Discussion Series. Washington, D.C.: Federal Reserve Board.

Granger, C. W. J., and P. Newbold. 1986. *Forecasting economic time series.* New York: Academic.

Granger, C. W. J., and T. Teräsvirta. In press. *Modelling dynamic nonlinear economic relationships.* Oxford: Oxford University Press.

Johansen, S. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12:231–54.

Johansen, S., and K. Juselius. 1990. Maximum likelihood estimation and inference on cointegration with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52:169–210.

Lee, T.-H., H. White, and C. W. J. Granger. 1993. Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics* 56:269–90.

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

Teräsvirta, T. 1990a. Power properties of linearity tests for time series. Economics Discussion Paper no. 90–15. University of California, San Diego.

Teräsvirta, T. 1990b. Specification, estimation and evaluation of smooth transition autoregressive models. Economics Discussion Paper no. 90–39. University of California, San Diego.

Teräsvirta, T., and H. M. Anderson. 1992. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics* 7:S119–S136.

Teräsvirta, T., C.-F. Lin, and C. W. J. Granger. 1993. Power of the neural network linearity test, *Journal of Time Series Analysis* 14:205–20.

Tong, H. 1990. *Non-linear time series analysis: A dynamic system approach.* Oxford: Oxford University Press.

White, H. 1989. An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. In *Proceedings of the International Joint Conference on Neural Networks,* vol. 2. San Diego, Calif: SOS Printing.

Comment

Andrew Harvey

Data mining is always a problem in time-series modelling. As Granger, Teräsvirta, and Anderson point out, this problem is particularly acute with nonlinear models, and it is certainly very important to assess the performance of models out of sample as well as in sample. The range of nonlinear models is very wide, and it seems virtually impossible to work within a particular class. Indeed, I have not been very convinced by attempts to work within classes of models such as bilinear or state dependent. In order to be useful, nonlinear models should have some economic meaning and a reasonably straightforward interpretation. As a simple example, we might consider an unobserved components model, consisting of a stochastic trend and a stochastic cycle of the kind considered in Harvey (1985) but with the period changing according to whether the change in observations in the previous period, Δy_{t-1}, is negative or positive.

The model described at the end of the previous paragraph is a nonlinear univariate model. The paper considers models with explanatory variables, and an additional source of nonlinearity here comes from nonlinearity in the functional form. This then raises the question of what kind of nonlinearity it is that the tests are picking up. Another issue surrounding the tests is that, if several are applied, it becomes difficult to say much about the Type I error. If enough powers of lagged variables are put into the test statistics, one is probably bound to reject at some stage. This being the case, the comment in the conclusion that overfitting due to data mining "is less likely to occur when tests for linearity have first found evidence against it before model estimation" becomes less convincing.

I feel that the tests for cointegration are a bit tangential since there is no strong economic reason for GNP and the index of leading indicators to be cointegrated and the tests are not being used to find other variables that might go in a set of cointegrated variables. More generally, the whole autoregressive distributed lag framework is not ideal for the kind of exercise being carried out here. Equation (21), with its large number of parameters and arbitrary lag structure, is not particularly appealing. Coupled with the difficulty of specifying the switching mechanism, and the associated lag, I suspect that the equation may not be very robust.

Reference

Harvey, A. C. 1985. Trends and cycles in macroeconomic time series. *Journal of Business and Economic Statistics* 3:216–27.

Andrew Harvey is professor of econometrics at the London School of Economics.