



## Sir Clive W.J. Granger Memorial Special Issue on Econometrics

### Sir Clive Granger's Contributions to Nonlinear Time Series and Econometrics

Timo Teräsvirta<sup>1,2</sup>

<sup>1</sup> *CREATES, Aarhus University, Denmark*

<sup>2</sup> *C.A.S.E., Humboldt-Universität zu Berlin, Germany*

---

**Abstract.** Clive Granger had a wide range of research interests and has worked in a number of areas. In this work the focus is on his contributions to nonlinear time series models and modelling. Granger's contributions to a few other aspects of nonlinearity are reviewed as well.

**2010 Mathematics Subject Classifications:** 62M10;91B84

**Key Words and Phrases:** cointegration, nonlinearity, nonstationarity, testing linearity

---

## 1. Introduction

Sir Clive Granger has been active in many areas of time series econometrics and made a large number of important contributions. As [49] puts it: 'It can be argued that Granger's is one of the most successful research programmes in the history of econometrics, and will be a lasting contribution to our discipline.' The introduction to the two volumes of Granger's collected papers, see [17], gives an idea of the scope and width of this programme, and the two volumes themselves contain a sizable amount of his published articles. The papers in this issue complement the picture.

In this paper the focus will be on Granger's work on nonlinear models and modelling. For space reasons, the exposition is restricted to conditional mean models, and Granger's contributions to modelling the conditional variance are not considered here. This review is necessarily rather brief, and more information can be found in the two books Granger has co-authored, [45] and [73]. The survey article [72] may also be helpful.

The paper is organised as follows. Granger's views on nonlinearity and building nonlinear models are presented in Section 2. Section 3 reviews his contributions to testing

---

*Email address:* tterasvirta@econ.au.dk (Timo Teräsvirta)k

linearity, and Section 4 considers nonlinear models Granger has developed and worked on. Section 5 presents Granger's ideas concerning nonstationary models and cointegration. In Section 6 the focus is on two cases in which nonlinear models may be confused with autoregressive models with long memory. Granger's work on forecasting with nonlinear models is presented in Section 7, and Section 8 contains short discussions on his contributions to a number of other 'nonlinear topics', including chaos. Final remarks can be found in Section 9.

## 2. Views

Before discussing Clive Granger's contributions to nonlinear econometrics, it may be instructive to consider his views on nonlinearity, in macroeconomic time series in particular. In this review I shall highlight three articles that contain such views. First, there is [26], based on an invited lecture at the Australasian meeting of the Econometric Society in 1992. The second is [29], a leading article in a conference issue of *Macroeconomic Dynamics*. Finally, [30] is based on a keynote address in a conference on nonlinear models and modelling. As already mentioned, Granger's views on chaos are presented in Section 8.

[26] is mostly concerned with building nonlinear models. He notes that the number of nonlinear models is large and that it may be difficult to choose among them. He lists a few families of such models, including the Fourier flexible form, the single hidden layer neural network model, projection pursuit, and the smooth transition regression model. He considers ways of selecting a parsimonious model within a given family, for example that of neural network models, using model selection criteria such as

$$C(m) = \ln \sigma^2(m) + m^d \frac{\ln T}{T} \quad (1)$$

where  $\sigma^2(m)$  is the error variance of a model with  $m$  parameters and  $T$  is the number of observations. When  $d = 1$ , (1) is the criterion (BIC) suggested by [66] and [67]. To keep the number of parameters low, Granger recommends using a 'super-parsimonious' version in which  $d > 1$ .

In probably the most important part of the paper Granger outlines a strategy for building nonlinear models for economic time series. His advice is to start with a small set of most relevant explanatory variables, perhaps one or two. Assuming the nonlinear family of models considered nests a linear model, the first step is to test linearity using a couple of different tests. Granger's contributions in this area are reviewed in Section 3. As already discussed, models to be fitted if linearity is rejected should be parsimonious.

The fitted model has to be evaluated using out-of-sample data. Granger's recommendation, not always followed by econometricians, is to save at least 20% of the sample for out-of-sample forecasting. He would like to compare one-step-ahead forecasts from the nonlinear model with ones from a (parsimonious) linear one. (Out-of-sample plays an important role in Granger's work: recall his opinion that the Granger noncausality hypothesis should be tested using the out-of-sample period; see [2].) In general, he would

accept the nonlinear model only if evidence favouring it is what he calls clear-cut. Finally, Granger stresses the fact that this strategy is applicable when the series are stationary. More discussion can be found in Granger and [73], Chapter 9, and [73], Chapter 16, where the strategy is applied to particular families of nonlinear models.

[29] opens with the statement that nonlinearity in macroeconomic time series is weak. Aggregation is claimed to be one of the reasons for this. There is more about the effect of aggregation on nonlinearity in Section 8.2. Interestingly, Granger also mentions seasonal adjustments as a potential cause for this weakness. He thinks, however, that nonlinearity may be stronger in multivariate contexts. As an example he takes the study by [68]. The authors analysed 214 US macroeconomic time series and found that nonlinear models fitted to them, albeit without first testing linearity, generated less accurate forecasts than a simple linear autoregressive model of order four. However, when the forecasts from nonlinear models were combined using equal weights, the performance of the nonlinear models improved, and the combined forecast was among the best ones in a majority of series.

Granger highlights nonlinear state space models as a promising tool and notes that they have not been much applied in practice. More about them can be found in [73], Chapter 9. He regards nonlinear error correction models as a success. They are discussed in Section 5. Finally, he points out that nonlinear models are inclined to overfit the data. One way of reducing the risk of this happening is testing linearity before fitting any nonlinear model. This was already emphasized in [26].

[30] discusses the possibility of moving from nonlinear models to models with time-varying parameters. As an argument he uses the result he calls White's Theorem because Hal White proved it for him. The theorem states that any nonlinear model can be viewed as a time-varying parameter (TVP) model. Let the  $y_t$  be an arbitrary time series such that  $E y_t < \infty$  and  $\Pr\{y_t = 0\} = 0$ . Then there exist sequences  $\{p_t\}$  and  $\{e_t\}$  such that  $p_t$  is  $\mathcal{F}_{t-1}$ -measurable where  $\mathcal{F}_{t-1}$  is the  $\sigma$ -field  $\sigma(\dots, y_{t-2}, y_{t-1})$ ,  $(e_t, \mathcal{F}_t)$  is a martingale difference sequence, and

$$y_t = p_t y_{t-1} + e_t.$$

Define

$$e_t = y_t - E\{y_t | \mathcal{F}_{t-1}\}$$

so the model for  $y_t$  is

$$\begin{aligned} y_t &= E\{y_t | \mathcal{F}_{t-1}\} + e_t = [E\{y_t | \mathcal{F}_{t-1}\} / y_{t-1}] y_{t-1} + e_t \\ &= p_t y_{t-1} + e_t \end{aligned}$$

where  $p_t = E\{y_t | \mathcal{F}_{t-1}\} / y_{t-1}$  is the TVP. Since  $E\{y_t | \mathcal{F}_{t-1}\}$  is not specified, Granger argues that the TVP can be 'a deterministic function of time, a specific function of a stochastic process such as a lagged observed time series or an unobserved series such as the common factor in the Kalman Filter'. In fact, this definition incorporates many models that are normally thought of as nonlinear time series models such as smooth transition, threshold autoregressive or Markov-switching autoregressive models. It also contains random coefficient models that are favoured by macroeconomists, at least if the coefficients are random

walks and the models thus nonstationary. All these models are discussed in [73], Chapter 3.

### 3. Testing linearity

Linear models have for a long time dominated the modelling of economic time series. Many nonlinear models that are applied in economic and econometric research nest a linear model. Besides, several of these models are only identified under the alternative, that is, when the nonlinear model is true. As Granger argues, see for example [26], it is therefore advisable to test linearity before fitting such a nonlinear model to the data.

Clive Granger and co-authors have made two contributions to this area of research. In [53] the aim is to compare a new test based on neural networks and originally developed by [77] with a large number of other well-known linearity tests. The model to be tested is

$$y_t = \mathbf{x}_t' \boldsymbol{\theta} + \sum_{j=1}^q \beta_j \psi(\mathbf{x}_t' \boldsymbol{\gamma}_j) + \varepsilon_t \quad (2)$$

where  $\{\varepsilon_t\}$  is a white noise process with zero mean and  $\psi(\cdot)$  is a logistic function, often called a 'hidden unit'. The weighted sum of the  $q$  hidden units is taken to be an approximation to the unknown form of nonlinearity present under the alternative. The null hypothesis, as formulated in the paper, is  $\Pr\{E(y_t|\mathbf{x}_t) = \mathbf{x}_t' \boldsymbol{\theta}\} = 1$  or,  $\beta_1 = \dots = \beta_q = 0$ , whereas the alternative is  $\Pr\{E(y_t|\mathbf{x}_t) = \mathbf{x}_t' \boldsymbol{\theta}\} < 1$ . It is seen that the model is not identified under  $H_0$  as  $\boldsymbol{\gamma}_j$ ,  $j = 1, \dots, q$ , are unidentified nuisance parameter vectors.

This identification problem was first studied by [11]. For more discussion, see [73], Chapter 5. The novelty in [53] is that the problem is solved by giving parameter vectors  $\boldsymbol{\gamma}_j$  in (2) values drawn randomly from a given distribution, so  $\psi(\mathbf{x}_t' \boldsymbol{\gamma}_j)$ ,  $j = 1, \dots, q$ , become observable random variables. A fairly large  $q$  is required for this Lee, White and Granger or LWG test to have reasonable power, but then this may lead to problems with nearly collinear regressors. As a remedy, the authors suggest principal components and selecting the most informative ones to replace the original hidden units. If the model is dynamic ( $\mathbf{x}_t$  contains lags of  $y_t$ ), the resulting test has an asymptotic  $\chi^2$ -distribution when the null hypothesis holds.

The power of this and a number of other tests is studied by simulation. The set of models contains both univariate and multivariate nonlinear models, including the bilinear model in Section 4.1 and the sign-autoregressive test in Section 3. The empirical size of the tests is investigated by having a first-order autoregressive model as the null model. No test turns out to dominate the others. However, the authors conclude that 'the new neural network test [...] appears to perform as well as or better than standard tests in certain contexts.' It may be mentioned that later on, [70] was able to explain many of the simulation results in [53] using analytic tools.

The starting-point of [71] is the same (but univariate) model. The identification problem is solved differently by approximating  $\psi(\mathbf{x}_t' \boldsymbol{\gamma}_j)$  by a third-order polynomial of

$\mathbf{x}_t = (1, y_{t-1}, \dots, y_{t-p})'$ . This leads to the following approximation:

$$y_t = \mathbf{x}_t' \theta^* + \sum_{i=1}^p \sum_{j \leq i} \delta_{ij} y_{t-i} y_{t-j} + \sum_{i=1}^p \sum_{j \leq i} \sum_{k \leq j} \delta_{ijk} y_{t-i} y_{t-j} y_{t-k} + \varepsilon_t^* \quad (3)$$

and the null hypothesis of this auxiliary model becomes  $H_0: \delta_{ij} = 0, i = 1, \dots, p, j = 1, \dots, i$ , and  $\delta_{ijk} = 0, i = 1, \dots, p, j = 1, \dots, i, k = 1, \dots, j$ . Under appropriate assumptions, this Teräsvirta, Lee and Granger or TLG statistic also has an asymptotic  $\chi^2$ -distribution when the null hypothesis is valid. Ignoring the third-order terms in (3) yields the well-known linearity test by [76].

Simulations suggest that the TLG test often has a better power than the LWG one. This, however, depends for instance on how many principal components is chosen for the LWG test. (Principal components can be used in connection with the TLG test as well.) In simulations the number was the same (two) as the one in the simulations of [53]. The authors point out that the asymptotic theory is not a reliable guide in small samples when the number of observations is small in relation to the dimension of the null hypothesis. They suggest using the F-distribution instead. [5] extended the test to work in the presence of conditional heteroskedasticity in errors.

At this point it may be interesting to consider [37] who write that model selection or information criteria should be preferred to statistical tests in economic modelling. If, however, the selection process involves nested models and the aforementioned identification problem is present, this suggestion becomes difficult to follow. As an example, consider the following smooth transition autoregressive (STAR) model ([73], Chapter 3):

$$y_t = \phi_0 + \phi_1 y_{t-1} + (\psi_0 + \psi_1 y_{t-1})(1 + \exp\{-\gamma(y_{t-1} - c)\})^{-1} + \varepsilon_t \quad (4)$$

where  $\varepsilon_t$  is zero mean white noise, and  $\gamma > 0$ . Equation (4) nests a linear model and becomes linear when  $\gamma = 0$ . The difference in the number of parameters between the STAR model and the linear autoregressive model to be used in the model selection criterion would thus equal one. However, (4) becomes linear also when  $\psi_0 = \psi_1 = 0$ , which suggests that this difference equals two. Sometimes it is taken to equal four because the nonlinear component contains four parameters. Because of the identification problem, the 'correct' number and thus the 'true' penalty remain unknown.

In this example, the use of model selection criteria leads to another problem. Suppose the nested linear model is true, and an information criterion such as AIC or BIC is used to select the model. This means estimating unidentified nonlinear models, which causes numerical problems as the estimation algorithm may not converge.

There exists, however, a way round this difficulty. The idea of [78] is to transform the nonlinear specification and estimation problem into a linear model selection problem. White applies it to building artificial neural network models that can nest a linear model, but it can also be used for constructing STAR models. For details, see [78], and for additional applications, [50] and [51].

## 4. Nonlinear models

### 4.1. Bilinear model

The first nonlinear model that has attracted Clive Granger's interest is the bilinear model examined at length in [31]. In this monograph the authors also discuss nonlinearity in general and mention another nonlinear model, the random coefficient autoregressive model

$$y_t + \alpha_{1t}y_{t-1} + \dots + \alpha_{pt}y_{t-p} = \varepsilon_t \quad (5)$$

where  $\alpha_{jt}$ ,  $j = 1, \dots, p$ , and  $\varepsilon_t$  are 'time-varying functions whose values are known at time  $t$ .' They, however, see equation (5) mainly as a difference equation whose solutions, after setting  $\varepsilon_t = 0$ , are the main object of interest.

Granger and Andersen distinguish between two main types of bilinear models: the Bilinear Autoregressive Moving Average (BARMA) model defined as

$$y_t = \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{k=0}^Q \sum_{\ell=1}^P \beta_{k\ell} \varepsilon_{t-k} y_{t-\ell} + \varepsilon_t \quad (6)$$

where  $\varepsilon_t$  is zero mean white noise, and the complete bilinear model is obtained by setting  $p = q = 0$  in (6). The BARMA( $p, q, P, Q$ ) model nests the linear AR model, and testing  $H_0: \beta_{k\ell} = 0, k = 1, \dots, Q; \ell = 1, \dots, P$ , amounts to testing linearity. Several special cases of the complete bilinear model are considered in the monograph.

The authors provide some examples of bilinear models in economics. One of them has to do with modelling stock returns with an MA(1) model. Defining returns as  $r_T = (y_t - y_{t-1})/y_{t-1}$ , where  $y_t$  is the price, one obtains

$$y_t = y_{t-1} + \varepsilon_t y_{t-1} + \theta_1 \varepsilon_{t-1} y_{t-1}.$$

At this point Clive Granger may not yet have known that in the near future he would have much to do with nonstationary models, albeit not with bilinear ones.

Much of the discussion in [31] focusses on finding stationary solutions for bilinear models, deriving moments of bilinear variables and conditions for invertibility. In particular, the authors show that  $\{\varepsilon_t^2\}$  is an autocorrelated process. This has prompted comparisons with ARCH models for the conditional variance; see for example [6]. Consequently, in simulations [53] report, the test of no ARCH by [62] performs well when the data-generating process is a bilinear model.

Invertibility conditions receive plenty of attention in the monograph as invertibility of bilinear models is crucial from the point of view of estimation and forecasting. The authors show how some simple nonlinear moving average models are non-invertible. They also derive invertibility conditions for diagonal bilinear models in which the bilinear component is of the form  $\beta_{kk} \varepsilon_{t-k} y_{t-k}$ . The considerations are extended to nondiagonal models in which the bilinear term equals  $\beta_{k, k+K} \varepsilon_{t-k} y_{t-k-K}$ ,  $K > 0$ .

[32] is devoted to the diagonal case with  $k = 1$ . For this model they derive the sufficient invertibility condition  $E \ln(\theta_1^2 y_t^2) = \ln \theta_1^2 + E \ln y_t^2 < 0$ . This and other conditions the

authors derived have generated plenty of interest. [64] showed that the sufficient condition for a more general model

$$y_t = \varepsilon_t + \theta_{kl}\varepsilon_{t-k}y_{t-\ell}, \quad k, \ell > 0$$

is  $\ln |\theta_1| + \mathbf{E} \ln |y_t| < 0$ , whereas the necessary condition equals  $\ln |\theta_1| + \mathbf{E} \ln |y_t| \leq 0$ .

Bilinear models also play a role in [21], where the topic is forecasting white noise. Contrary to common beliefs, Granger demonstrates that white noise can indeed be forecastable. The paper contains several examples, the simplest one being the bilinear model

$$y_t = \theta_{12}\varepsilon_{t-1}y_{t-2} + \varepsilon_t \quad (7)$$

where  $\varepsilon_t$  is what Granger calls pure white noise:  $\varepsilon_t$  and  $\varepsilon_{t-k}$  are independent for all  $t$  and  $k$ . Gaussian white noise is an example of pure white noise. If  $\varepsilon_t$  is pure white noise in (7), then  $y_t$  is white noise as  $\mathbf{E}y_t = 0$  and  $\mathbf{E}y_t y_{t-k} = 0$  for  $k \neq 0$ . The model (7) is invertible if  $\theta_{12}\sigma_\varepsilon < 1/\sqrt{2}$ , and the coefficient of determination equals  $R^2 = \theta_{12}^2\sigma_\varepsilon^2$ , where  $\sigma_\varepsilon^2 = \mathbf{E}\varepsilon_t^2 < \infty$ . Thus, when  $\theta_{12}\sigma_\varepsilon \rightarrow 1/\sqrt{2}$ ,  $R^2 \rightarrow 1/2$ , so there is plenty of (nonlinear) forecastability in the model. Granger issues a warning: when one estimates a model such that the errors (judging from the residuals) appear to be white noise, they may still be forecastable.

Research on bilinear models has generated a host of theoretical results concerning probabilistic properties of the model and properties of estimators; see, for example, [69]. Nevertheless, the model has not turned out to be very useful in economic research. The reason seems to be that it mostly describes atypical behaviour that manifests itself in the form of outliers. See [59] for an illuminating discussion. Some other nonlinear models such as the threshold autoregressive, the smooth transition autoregressive or the hidden Markov autoregressive model have been more successful in this respect. They are discussed in [73], Chapters 3 and 16.

Granger and Andersen also consider building bilinear models. They mention the three-stage model building strategy of [7], in which the linear autoregressive-moving average model is first specified ('identification'), the specified model estimated, and the estimated model evaluated ('diagnostic checking'). They write that the same strategy, suitably modified, should be used in constructing bilinear models. Granger's contributions to an important part of specification of nonlinear models, namely testing linearity, was already considered in the previous section.

## 4.2. The m-m model

As discussed in Section 2, Granger's view has been that in practice, nonlinearity in macroeconomic series is often weak. In introducing the m-m process in [35] the authors point out that while this is the case in univariate models, the evidence of nonlinearity becomes stronger when a bivariate system is considered. They define a bivariate nonlinear model called the m-m model as follows:

$$x_{t+1} = \max(\alpha x_t + a, \beta y_t + b) + \varepsilon_{x,t+1} \quad (8)$$

$$y_{t+1} = \min(\gamma x_t + c, \delta y_t + d) + \varepsilon_{y,t+1} \quad (9)$$

where  $\varepsilon_{x,t+1}$  and  $\varepsilon_{y,t+1}$  are independent and identically distributed with mean zero and mutually independent variances  $\sigma_x^2$  and  $\sigma_y^2$ , respectively. The 'm-m' in the title of the paper indicates that the process may also be defined as max-max or min-min. As the equations (8) and (9) suggest, this is a highly nonlinear system that does not nest a linear model. When  $\alpha = \beta = \gamma = \delta = 1$ , the resulting system is an integrated m-m process. Setting  $a = b = c = d = 0$  one obtains what is called a stationary m-m process. The model is similar to the linear vector autoregressive model in the sense that neither variable is exogenous. It is seen from the equations (8) and (9) that testing the Granger noncausality hypothesis, see Hendry, this issue, is possible in this framework.

[35] consider equilibrium values  $x$  and  $y$  of the system. These are values obtained from the model when  $t \rightarrow \infty$ , assuming  $\sigma_x^2 = \sigma_y^2 = 0$  for all  $t > T_0 > 0$ . Whether or not they exist depends on the parameters of the system, and the appropriate conditions are derived in the article. Special attention is given to the equilibrium values of  $z_t = x_t - y_t$ . The paper contains a proposition that generally  $\{x_{t+1}\}$  and  $\{y_{t+1}\}$  are nonlinearly integrated processes, that is, their variance becomes infinite with the number of observations. There do exist parameter restrictions that make  $z_t = x_t - y_t$  stationary and ergodic, which means that the nonlinear variables  $x_t$  and  $y_t$  are linearly cointegrated. It turns out that the equilibrium conditions for  $z_t$  are necessary for this variable to be stationary and ergodic. There is also some discussion of more general forms of linear cointegration than  $(1, -1)$ . Further, Granger and Hyung argue that because the cointegration relationship is linear, tests of cointegration derived for standard unit root processes work for m-m variables, although  $x_t$  and  $y_t$  are highly nonlinear.

Estimation of m-m processes with maximum likelihood is not possible because of discontinuities in the equations (8) and (9). [35] suggest the following approximation:

$$x_t = \frac{1}{s} \log(e^{s(\alpha x_t + a)} + e^{s(\beta y_t + b)}) + \varepsilon_{x,t+1}^* \quad (10)$$

$$y_t = -\frac{1}{s} \log(e^{s(\gamma x_t + c)} + e^{s(\delta y_t + d)}) + \varepsilon_{y,t+1}^*. \quad (11)$$

The argument in favour of using the equations (10) and (11) is that when the scale parameter  $s \rightarrow \infty$ , this system converges to the m-m process defined by the equations (8) and (9). The likelihood is well behaved for finite values of  $s$  and choosing a large value for it, maximum likelihood estimation of the parameters of an approximate m-m system is possible. One could also apply derivative-free methods such as simulated annealing or genetic algorithms; see for example [73], Chapter 12, but the authors do not consider that possibility.

[35] conduct a simulation study to find out how well standard linearity tests respond to nonlinearity of m-m type. The design of the experiment closely resembles that in [53]. The same tests are used in both papers. Four m-m models are simulated, and the power of the tests varies greatly with the model. An interesting detail is that there is a stationary model in which even the [62] test has power, albeit less than the other tests. Since, as already mentioned, this is a test of no ARCH, it appears that some m-m models can generate series that contain conspicuous outliers or rather volatile periods.

Finally, the workings of the m-m model are illustrated using the interest rate of a six-month commercial paper and a three-month Treasury bill. The estimation period extends from January 1947 until December 1986 (480 observations) and the forecasting period from January 1987 to September 1997 (129 observations). In the estimated model, the monthly interest rates are differenced, and the T-bill rate equation is linear. The same battery of tests as before is conducted, and most of them reject linearity, including the test of [62]. Since monthly interest rate series may well contain conditional heteroskedasticity, this should not be surprising. The forecasts from the m-m model turn out to be more accurate than the ones from a linear vector autoregressive model and a threshold cointegration model by [3]. Despite this success, there are few applications of the m-m model. The reason may be that while very original, the structure of the model may not sufficiently resemble that of any economic theory model.

### 4.3. Momentum threshold autoregressive model

Unlike the m-m model, models with switches or thresholds have been quite popular in economic applications. In the econometric literature [18], Chapter 9, introduced a switching regression model with two regimes in which the observations are independent and switching is controlled by a strongly exogenous variable. [74] presented an autoregressive model he called the self-exciting threshold autoregressive (SETAR) model. [75] contains a comprehensive account of this model and its dynamic properties, hypothesis testing and parameter estimation are also discussed.

[15] look at the simple Dickey-Fuller autoregression

$$\Delta y_t = \rho y_{t-1} + \varepsilon_t$$

where  $\varepsilon_t$  is (nonforecastable) white noise. They note that the adjustment, which exists for  $-2 < \rho < 1$ , is symmetric around zero and argue that it could in fact be asymmetric. This leads them to consider the following TAR model:

$$\begin{aligned} \Delta y_t &= \rho_1 I(y_{t-1} \geq c) y_{t-1} + \rho_2 \{1 - I(y_{t-1} \geq c)\} y_{t-1} + \varepsilon_t \\ &= \rho_2 y_{t-1} + (\rho_1 - \rho_2) I(y_{t-1} \geq c) y_{t-1} + \varepsilon_t \end{aligned} \quad (12)$$

where  $I(y_{t-1} \geq c)$  is an indicator function:  $I(A) = 1$  when  $A$  is true, zero otherwise. Since the interest lies in asymmetry around zero, [15] assume  $c = 0$  in (12). This means that their model is in fact linear. The novelty, the momentum TAR (M-TAR) model, is obtained by replacing  $y_{t-1}$  in the indicator function by  $\Delta y_{t-1}$ . The resulting model is linear as well. Another version of the M-TAR one is obtained by demeaning, in which case (12) becomes

$$\begin{aligned} \Delta y_t &= \rho_1 I(y_{t-1} \geq \hat{\mu}_y) (y_{t-1} - \hat{\mu}_y) + \rho_2 \{1 - I(y_{t-1} \geq \hat{\mu}_y)\} (y_{t-1} - \hat{\mu}_y) + \varepsilon_t \\ &= \rho_2 y_{t-1} + (\rho_1 - \rho_2) I(y_{t-1} \geq \hat{\mu}_y) y_{t-1} + \varepsilon_t \end{aligned} \quad (13)$$

where  $\hat{\mu}_y$  is the sample mean. This model is still linear because  $\hat{\mu}_y$  is estimated prior to estimating the other parameters of the model.

The authors consider testing the unit root hypothesis  $\rho_1 = \rho_2 = 0$ . To this end, they simulate the null distribution of tests of this null hypothesis assuming both  $c = 0$  and  $c = \hat{\mu}_y$  in (12) and (13).

In this set-up, nonlinearity enters (12) through the back door. Granger and Enders remark that zero or the sample mean may be biased estimates of the threshold parameter  $c$ . If the null hypothesis is rejected, they suggest estimating this parameter by giving it sample values of  $y_{t-1}$  or  $\Delta y_{t-1}$  within a certain range that excludes the smallest and largest values and choosing the value that minimises the sum of squared errors. This is how  $c$  is generally estimated in TAR models. The model does not need not be of order one, that is, the regimes may have longer lags than one.

An advantage of this procedure is that the identification problem already mentioned in Section 3 arising when  $c$  is unknown is avoided. A disadvantage is that the test may not always reject the null model when the alternative holds but  $c$  deviates from either zero or  $\hat{\mu}_y$ , depending on whether one applies (12) or (13). One solution to this inconvenience can be found in [14], Chapter 1. His model has the following form

$$\Delta y_t = \phi_0 + \phi_1 \Delta y_{t-1} + \psi y_{t-1} + (\varphi_0 + \varphi_1 \Delta y_{t-1}) G(y_{t-1}; \gamma, c) + \varepsilon_t \quad (14)$$

where

$$G(y_{t-1}; \gamma, c) = (1 + \exp\{-\gamma(y_{t-1} - c)\})^{-1} - 1/2, \quad \gamma > 0 \quad (15)$$

is the logistic transition function; see for example [73], Chapter 3. This model is different from (12) in that it contains two free intercepts,  $\phi_0$  and  $\varphi_0$ , and a continuum of regimes. If, however,  $\phi_i = \varphi_i = 0$ ,  $i = 0, 1$ , and  $\gamma \rightarrow \infty$  in (15), (14) collapses into (12). A combined unit root and linearity test against (14) can be formulated as the joint test of  $\psi = 1$  and  $\gamma = 0$ , in which case  $\varphi_0$ ,  $\varphi_1$ , and  $c$  are unidentified when the null hypothesis holds.

An operational test statistic is derived following [58]. This implies approximating (14) by its first-order Taylor approximation around zero. Doing so yields the auxiliary model

$$y_t = \phi_0^* + \phi_1^* \Delta y_{t-1} + \lambda y_{t-1} \Delta y_{t-1} + \rho y_{t-1} + \varepsilon_t^* \quad (16)$$

where  $\varepsilon_t^* = \varepsilon_t$  when the new null hypothesis  $\lambda = 0$ ,  $\rho = 1$  is valid. Eklund derives the asymptotic null distribution of his F-statistic and simulates the relevant critical values. When  $\lambda = 0$ , (16) becomes the ADF test equation with an intercept and one lag. The alternative model is larger than in [15] because it covers all positive and finite values of  $\gamma$ . Strictly speaking, the asymptotic theory is not valid for  $\gamma = \infty$ , but the test has reasonable power against this alternative as well. Comparing this approach with the one in [15] could be interesting.

## 5. Nonstationarity and nonlinearity

### 5.1. Nonlinear transformations of nonstationary processes

What happens to nonstationary variables, for instance ones with a unit root, when they are transformed using nonlinear transformations? How would the probabilistic linear

properties such as the mean, variance, and the autocovariance structure in particular, of the transformed variables look like? [16] study these questions and provide answers. They consider polynomial, exponential and periodic transformations as well as the probability integral transform of

$$x_t = \mu + mt + \sum_{j=1}^t \varepsilon_j, \quad j = 1, \dots, t$$

where  $\mu$  is the starting-value of the process at  $t = 0$ , and  $\varepsilon_t \sim \text{iid}\mathcal{N}(0, \sigma_\varepsilon^2)$ . It is seen that  $x_t$  is I(1) with drift. Under these assumptions,  $\mathbb{E}x_t = \mu + mt$  and  $\text{cov}(x_t, x_{t-k}) = (t-k)\sigma_\varepsilon^2$  for  $|k| \leq t-1$ . Since  $\varepsilon_t$  is normal,  $x_t \sim \mathcal{N}(\mu + mt, \text{var}(x_t))$ , where  $\text{var}(x_t) = t\sigma_\varepsilon^2$ . Define the standard normal variable  $z_t = (x_t - \mu - mt)/\sqrt{\text{var}(x_t)}$ . The transformation  $y_t = \mathbb{T}(x_t)$  is approximated by Hermite polynomials that are defined by standard normal variables:

$$y_t = \mathbb{T}(\mu + mt + \sqrt{\text{var}(x_t)}z_t) = \sum_{i=0}^M \alpha_j^{(t)} H_j(z_t)$$

where

$$\alpha_j^{(t)} = \frac{1}{j!} \mathbb{E} \left( \frac{d^j \mathbb{T}(\mu + mt + \sqrt{\text{var}(x_t)}z_t)}{dz_t^j} \right). \tag{17}$$

It is not possible to review all results of [16] here. The focus will be on selected outcomes on the exponential and the probability integral transformation. Consider first the exponential transformation

$$y_t = \exp(\mu + mt + \sqrt{\text{var}(x_t)}z_t) = \sum_{i=0}^M \alpha_j^{(t)} H_j(z_t)$$

where, from (17),

$$\alpha_j^{(t)} = \exp(\mu + mt + \frac{1}{2} \text{var}(x_t)z_t) \frac{\text{var}(x_t)^{j/2}}{j!}.$$

It turns out that  $\mathbb{E}y_t = \exp\{\mu + \theta t\}$  where  $\theta = m + \sigma_\varepsilon^2/2$ . Setting  $\theta = 0$  makes the expectation constant. Autocorrelations of  $y_t$ , if they are nevertheless well defined using the standard definition, become

$$\text{corr}(y_t, y_{t-k}) = \exp\{-k\sigma_\varepsilon^2/2\}$$

and so decay exponentially as in the stationary AR(1) model. Differencing yields  $\mathbb{E}\Delta y_t = (1 - \exp\{-\theta\}) \exp\{\mu + \theta t\}$  which is not constant either unless  $\theta = 0$ . Consequently, the transformed variable  $y_t$  cannot be made stationary by differencing.

The probability integral transform of  $x_t$  that [16] called the neural network transformation, is considered in the case  $\mu = m = 0$ . Then, assuming normality,

$$y_t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_t} \exp\{-\frac{s^2}{2}\} ds. \tag{18}$$

Due to the nature of the transformation, the authors find that  $y_t$  still has the long memory property. In order to consider the case  $m \neq 0$ , a neural network transformation should involve a linear combination of normal distributions with different means and variances, for example

$$y_t = \beta_0 + \sum_{j=1}^q \frac{\beta_j}{\sigma_j \sqrt{2\pi}} \int_{-\infty}^{x_t} \exp\left\{-\frac{(s - \mu_j)^2}{2\sigma_j^2}\right\} ds.$$

Since Granger and Ermini want to keep the transformations as simple as possible, they do not discuss this extension.

[28] continues this discussion by using the concept of balance. To illustrate this, consider two random variables  $y_t$  and  $x_t$  and an equation

$$y_t = g(x_t) + \varepsilon_t \quad (19)$$

where  $\varepsilon_t$  is white noise. The equation (19) is balanced if the left-hand and right-hand side possess the same properties. For example, assume that  $y_t$  and  $x_t$  are random walks. Then  $g(x_t)$  also has to be a random walk, otherwise (19) is not a balanced equation. This requirement excludes a large number of nonlinear transformations. One message of this is that when economic theorists write up models linking economic variables, they have to be aware of their properties and those of the transformations so that they do not end up with unbalanced equations. More generally, [28] contains a lot of interesting material that for space reasons cannot be discussed here.

It may be added that [13] examine transformations of fractionally integrated processes. A process  $\{x_t\}$  is called fractionally integrated with the memory index  $d$ ,  $I(d)$  for short, if

$$x_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$$

where  $c_j = \Gamma(j+d)/\{\Gamma(j+1)\Gamma(d)\}$ . Only the case where  $x_t$  is nonstationary and, more specifically,  $1/2 < d < 1$ , is considered here. The authors show that in this case, power transformations of  $x_t$  are also nonstationary, and the memory index remains unchanged. For the square transformation,  $y_t = x_t^2$ , they derive the result analytically. For higher order transformations it is obtained by simulation.

Dittmann and Granger also study the logistic transformation which is very similar to (18), as

$$y_t = (1 + \exp\{-x_t\})^{-1}.$$

When  $x_t \sim I(d)$ ,  $1/2 < d < 1$ , they find that  $y_t \sim I(d_1)$ , where  $d_1 < d$ , but nevertheless  $1/2 < d_1 < 1$ , so  $y_t$  is still nonstationary. As the authors point out,  $y_t$  is bounded. The logistic transformation thus shows that a nonstationary fractionally integrated process can be bounded.

Granger's work on nonlinear transformations of fractionally integrated processes may prompt one to ask: has there been cross-fertilisation? Has his work on nonlinear models somehow aroused his interest in fractional integration or vice versa? The latter is hardly

true as Granger had already studied bilinear models before considering fractional integration. In [19], he shows how long memory may arise from aggregating *linear* autoregressive (but not moving average) models. He does not mention nonlinearity in that context. The two areas touch each other later, see Section 6.

## 5.2. Stochastic unit root processes

Processes in which nonstationarity is due to a unit root have been extensively studied in time series and econometrics literature. [44] introduce a generalisation in which the unit root is stochastic. More precisely, they assume that the random variable  $y_t$  is generated by

$$y_t = \exp\{\alpha_t\}y_{t-1} + \varepsilon_t \quad (20)$$

where  $\alpha_t$  has the form

$$\alpha_t = \mu + \rho\alpha_{t-1} + \eta_t \quad (21)$$

with  $|\rho| < 1$  and  $\eta_t \sim \text{iid}\mathcal{N}(0, \sigma_\eta^2)$  and independent of  $\varepsilon_t \sim \text{iid}(0, \sigma^2)$ . Denoting  $E\alpha_t = \mu/(1 - \rho) = m$ , one obtains the following condition for the mean of the time-varying coefficient to equal one:  $m + \sigma_\alpha^2/2 = 0$ , where  $\sigma_\alpha^2 = \sigma_\eta^2/(1 - \rho^2)$ . This follows from the properties of the normal distribution ( $\eta_t$  was assumed normal), and the subsequent model is called STUR A. This outcome may be compared with the results of [16] in the previous subsection, where an exponentially transformed random variable has a constant mean when this restriction holds.

The stochastic root fluctuates around unity, which means that the STUR process defined by equations (20) and (21) may contain explosive as well stationary periods. How persistent they are depends on  $\rho$  and  $\sigma_\eta^2$  in (21). [44] define another STUR model called STUR B, but it is ignored here. Instead of  $\exp\{\alpha_t\}$  and (21), [55], in a parallel development, write (20) as  $y_t = \alpha_t y_{t-1} + \varepsilon_t$ , where  $\alpha_t = 1 + \delta_t$  with  $\delta_t = \rho\delta_{t-1} + \eta_t$ ,  $|\rho| \leq 1$ . [61] and [81] consider a special case of this, where  $\rho = 0$ .

The randomness of  $\alpha_t$  makes STUR models quite different from the standard unit root process, obtained in STUR A by setting  $\sigma_\eta^2 = m = 0$ . Obviously, if  $\sigma_\eta^2$  is small, stochastic unit roots are in practice quite difficult to distinguish from standard unit roots. The usual 'power failures' associated with the ADF unit root test should therefore be apparent when the unit root is stochastic as in (20) and (21). The simulations run in [44] support this conclusion. As [54] and, later, [81] point out, the STUR process becomes very erratic when  $\sigma_\eta^2$  increases. Parameter estimation by maximum likelihood is complicated by the fact that  $\alpha_t$  is unobserved, and [44] therefore maximise the likelihood by simulation. This involves independent draws from distributions of  $\alpha_t$ .

Another property of the stochastic unit root process is that, as observed by [54], it cannot be made weakly stationary by differencing. This is also discussed in [16] in connection with the exponential transformation. The original process is strictly stationary, but the marginal distribution of  $y_t$  has no finite moments. This and other properties of STUR A are studied by [80]. As the author points out, sample autocorrelations of the STUR process therefore have no meaning. Yoon's paper also contains references to other

researchers' work on STUR. More discussion on STUR processes and further references can be found in [73], Chapter 11.

### 5.3. Nonlinear cointegration

As Clive Granger developed the concept of (linear) cointegration discussed in this issue by Castle and Hendry, it is only natural that he would be interested in nonlinearities in nonstationary series and nonlinear cointegration. The first time he turns to this extension is in [22], see also [38]. He points out that the fact that a set of economic variables are not linearly cointegrated 'does not necessarily imply that there is no equilibrium relationship among them.' In order to study this possibility further, Granger introduces time-varying parameter (TVP) cointegration. In doing so, he first defines a univariate TVP-I(0) process. This is done by first using the result of [10] stating that any univariate process  $\{y_t\}$  with  $Ey_t = 0$  and bounded variance for all  $t$  has a generalised Wold representation

$$y_t = \sum_{j=0}^{\infty} c_{jt} \varepsilon_{t-j}$$

where  $c_{jt}$  is a deterministic double (indexed by  $j$  and  $t$ ) sequence such that  $\sum_{j=0}^{\infty} c_{jt}^2 \leq M < \infty$  for all  $t$ , and  $\varepsilon_t$  is zero mean white noise. The time-varying parameter I(0) process is defined thereafter through the evolutionary spectrum

$$f_t(\omega) = \frac{1}{2\pi} \left| \sum_{j=0}^{\infty} c_{jt} \exp\{i\omega j\} \right|^2$$

see [63], as follows: If a time series  $y_t$  with no deterministic component has an evolutionary spectrum  $f_t(\omega)$  that is bounded from above and positive for all  $t$  and  $\omega$ , then  $\{y_t\}$  is TVP-I(0). (Recall that [20] defines cointegration using the frequency domain.)

The linear I(1) process can be generalised to the TVP-I(1) process in the same way. Let  $\mathbf{y}_t$  be an  $N$ -vector such that

$$\mathbf{y}_t = \mathbf{C}_t(\mathbf{B})\mathbf{w}_t \tag{22}$$

where  $\mathbf{B}$  is the lag operator,  $\mathbf{w}_t = \sum_{j=1}^t \boldsymbol{\varepsilon}_{t-j}$  with  $\boldsymbol{\varepsilon}_t$  being vector white noise, and  $\mathbf{C}_t(0) = \mathbf{I}_N$  for all  $t$ . Furthermore, defining  $\mathbf{C}_t(z) = \sum_{j=0}^{\infty} \mathbf{C}_{jt} z^j$ , assume  $\sum_{j=0}^{\infty} \mathbf{C}_{jt} \boldsymbol{\Omega} \mathbf{C}'_{jt} < \infty$  for all  $t$ , where  $\boldsymbol{\Omega} = E\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t$ . If these conditions are satisfied, then  $\{\mathbf{y}_t\}$  is TVP-I(1).

Cointegration can now be defined in the usual way. Write

$$\mathbf{C}_t(\mathbf{B}) = \mathbf{C}_t(1) + \mathbf{C}_t^*(\mathbf{B})(1 - \mathbf{B})$$

and rewrite (22) as

$$\mathbf{y}_t = \mathbf{C}_t(1)\mathbf{w}_t + \mathbf{C}_t^*(\mathbf{B})\boldsymbol{\varepsilon}_t$$

TVP cointegration occurs if there exists an  $N$ -vector  $\boldsymbol{\alpha}_t$  such that  $\boldsymbol{\alpha}'_t \mathbf{C}_t(1) = \mathbf{0}'$  for all  $t$ . This is no doubt a strong condition.

It may be mentioned that [39] introduce non-symmetry or asymmetry into an error-correction model of the relationship between sales  $s_t$  and production  $p_t$  that involves positive and negative inventories. This is done by splitting a stationary variable or a linear combination of nonstationary variables,  $x_t$ , say, to two components,  $x_t^+ = \max\{x_t, 0\}$  and  $x_t^- = \min\{x_t, 0\}$  and entering both in the same equation. Such a model allows asymmetry in the response of  $\Delta s_t$  and  $\Delta p_t$  to changes in  $x_t$ , but the model is still linear in parameters and will therefore not be studied here.

[33] examine the effect of transformations of integrated variables on unit root tests. The conclusion is that if the data are monotonically transformed, the unit root test applied should be a rank-based one as it is invariant to such transformations. [33] also ask the following question. Suppose  $x_t$  and  $y_t$  are two variables that are I(1) and cointegrated. Are the nonlinearly transformed variables  $g(x_t)$  and  $g(y_t)$  also cointegrated? To answer the question, they assume that

$$x_t = \mu + x_{t-1} + e_t$$

where  $e_t$  is white noise and independent of  $x_t$ , and that

$$y_t = \alpha x_t + \varepsilon_t$$

where  $\varepsilon_t$  is stationary. This means that  $y_t$  and  $x_t$  are linearly cointegrated. Then, by a mean-value expansion,

$$g(y_t) = g(\alpha x_t + \varepsilon_t) = g(\alpha x_t) + \varepsilon_t g'(\alpha x_t + r_t)$$

where  $r_t$  is a residual. Now,  $\varepsilon_t g'(\alpha x_t + r_t)$  is (mean) stationary with mean zero and some heteroskedasticity, so  $g(y_t) - g(\alpha x_t)$  is mean stationary. This implies that  $g(y_t)$  and  $g(x_t)$  are cointegrated if  $\alpha = 1$  or if  $g(\alpha x_t) = \alpha^k g(x_t)$ , in which case the cointegrating parameter is  $\alpha^k$ . Note that the cointegrating relationship considered is linear.

[27] briefly considers the case of nonlinear cointegration where there exist  $r$  linear cointegrating relationships between the elements of  $\mathbf{y}_t$ :  $\mathbf{z}_t = \mathbf{B}'\mathbf{y}_t$ . The error-correction model becomes

$$\Phi(\mathbf{B})\Delta\mathbf{y}_t = \Gamma\mathbf{f}(\mathbf{z}_{t-1}) + \varepsilon_t$$

where  $\mathbf{f}(\mathbf{z}_t) = (f_1(\mathbf{z}_t), \dots, f_r(\mathbf{z}_t))'$  a vector of nonlinear functions. Two examples for  $r = 1$  are given:  $f(z_t) = z_t^3$  and  $f(z_t) = \max(z_t, 0)$ , the latter as in [39]. Granger points out that the parameters in  $\mathbf{z}_t$  are usually estimated in the linear framework. He mentions that it is unclear whether that is efficient or not. This may still be an unresolved issue.

[43] look at many different types of nonlinear cointegration. To cite the authors, their aim is 'to suggest and examine generalizations [of cointegration] whilst maintaining the idea of cointegration and, consequently, to provide ways of making interpretations of the results of cointegration analysis both more realistic and more useful'. One of the generalisations is nonlinear cointegration. They illustrate it by the following bivariate example: Begin by generating a pair of possibly unobserved univariate series  $z_t$  and  $w_t$  from

$$z_t = \lambda_t z_{t-1} + \varepsilon_{zt}, \quad |\lambda_t| < 1 \tag{23}$$

$$w_t = \phi_t w_{t-1} + \varepsilon_{wt}, \quad |\phi_t| > 1 \tag{24}$$

where  $\{\varepsilon_{zt}\}$  and  $\{\varepsilon_{wt}\}$  are martingale difference sequences,  $z_t = x_t - a_t y_t$  and  $w_t = c_1 x_t + c_2 y_t$ . Note the time-varying parameters  $\lambda_t, \phi_t, a_t$  and  $c_{2t}$ . Assuming  $c_{2t} = 1 - c_1 a_t$  (there is a typo in the paper) yields the pair

$$x_t = c_{2t} z_t + a_t w_t \tag{25}$$

$$y_t = -c_1 z_t + w_t \tag{26}$$

so  $x_t$  and  $y_t$  are expressed as linear functions of  $z_t$  and  $w_t$ . Inserting (23) and (24) into (25) and (26) gives the following nonlinear error-correction model:

$$(1 - \phi_t \mathbf{B})x_t = \gamma_{1t} z_{t-1} + \varepsilon_{xt}$$

$$(1 - \phi_t \mathbf{B})y_t = \gamma_{2t} z_{t-1} + \varepsilon_{yt}$$

where  $\gamma_{1t} = c_{2t}(\lambda_t - \phi_t)$  and  $\gamma_{2t} = -c_1(\lambda_t - \phi_t)$ . The error term  $\varepsilon_{xt}$  equals

$$\varepsilon_{xt} = \phi_t(a_t - a_{t-1}) + c_{2t}\varepsilon_{zt} + a_t\varepsilon_{wt}.$$

It is seen that assuming  $\varepsilon_{xt}$  to be white noise requires  $a_t = a_{t-1}(= a)$ , that is, the cointegrating relationship is stable. Obviously, deriving an error-correction model with white noise errors in which  $a_t$  is time-varying is not completely straightforward if the equations (23) and (24) form the starting-point.

Granger and Swanson discuss possible representations for  $a_t$ . They suggest that it may be a stochastic unit root process or that it may change smoothly over time. The alternative that  $a_t$  is stochastic unit root process may not have been studied in the literature. Assuming  $a_t$  stochastic has been considered (in a broad sense) by [48]. They postulate the  $m$ -dimensional model

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{\Pi}_t \mathbf{w}_t + \boldsymbol{\varepsilon}_t$$

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \boldsymbol{\eta}_t$$

$$\mathbf{\Pi}_t = \mathbf{\Pi} + \mathbf{V}_t$$

where  $\boldsymbol{\varepsilon}_t, \boldsymbol{\eta}_t$  and  $\text{vec}(\mathbf{V}_t)$  are stationary vectors with mean zero. In this framework the authors define the concepts of stochastically integrated vector and stochastic cointegration. Asymptotic inference is discussed and the estimators simulated. No empirical example is given, however. Continuing this line of research, [60] derive a test of stochastic cointegration against no cointegration.

In [65], the long-run parameter changes smoothly over time. Using the bivariate model in [43] as an example, the error-correcting term  $z_t = y_t - a_t x_t$ , where  $a_t = G(t/T; \gamma, c)$ , is a deterministic function of (rescaled) time and  $T$  is the number of observations. Furthermore, as in [56],

$$G(t/T; \gamma, c) = (1 + \exp\{-\gamma(t/T - c)\})^{-1}, \quad \gamma > 0$$

or

$$G(t/T; \gamma, c) = 1 - \exp\{-\gamma(t/T - c)^2\}, \quad \gamma > 0.$$

The other parameters are assumed constant, although this assumption could be relaxed. Asymptotic inference and testing the null hypothesis  $a_t \equiv a$  are discussed. This paper also contains an empirical application.

Returning to [43], they also show how a certain myopic cost function leads to a nonlinear error-correction model. Assuming the elements of the  $N$ -vector  $\mathbf{y}_t$  are  $I(1)$  and that  $\mathbf{z}_t = \mathbf{A}'\mathbf{y}_t$ , where  $\mathbf{z}_t$  is an  $r \times 1$  vector,  $r < N$ ,  $\mathbf{A}$  is an  $N \times r$  matrix, has all its elements  $I(0)$ . This cost function is a nonlinear function  $\mathbf{z}_t$ . The agent is minimizing the cost of straying from the equilibrium defined as  $\mathbf{z}_t = \mathbf{0}$ . It turns out, after some algebra, that this leads to the following error-correction model:

$$\Delta \mathbf{y}_t = \mathbf{\Gamma} \mathbf{g}(\boldsymbol{\delta}' \mathbf{z}_{t-1}) + \boldsymbol{\varepsilon}_t \quad (27)$$

where  $\mathbf{g}(\boldsymbol{\delta}' \mathbf{z}_t) = (g_1(\boldsymbol{\delta}'_1 \mathbf{z}_t), \dots, g_r(\boldsymbol{\delta}'_r \mathbf{z}_t))'$ . Each equation of (27) contains  $r$  linear combinations of the  $I(0)$  variable  $\mathbf{z}_t$ , but they enter these equations nonlinearly. Since Granger and Swanson present ideas rather than applications, models of this type are neither specified nor estimated in the paper. They do have an example of the (bivariate) situation in which nonlinearity is such that the cointegration component  $g_1(z_t) = 0$  for  $|z_t| < z^*$ , whereas  $g_1(z_t) = \gamma(z - z^*)^\alpha$ ,  $\alpha > 0$ , with  $g_1(z) = g_1(-z)$ , outside this band. This means that the model does not 'equilibrium correct' when  $z_t$  is sufficiently close to the equilibrium point zero. This idea is similar to threshold cointegration as defined by [3]. For more discussion on nonlinear cointegration, see [73], Chapter 11, and references therein.

## 6. Nonlinear models with misleading linear properties

[16] conclude: 'The relevance of these results for econometric practice should not be underestimated: evaluation techniques, using autocorrelations and regressions of the future on the present, are all essentially linear and thus potentially misleading when considering nonlinearity.' Granger returns to this theme in the form of a simple example in [46]. The model is

$$y_t = \text{sgn}(y_{t-1}) + \varepsilon_t \quad (28)$$

where  $\varepsilon_t \sim \text{iid}\mathcal{N}(0, \sigma^2)$ , and the sign function  $\text{sgn}(x) = 1$ , if  $x > 0$ ,  $= -1$ , if  $x < 0$ , and zero otherwise. This sign-autoregressive model is also used in simulations of [53]. It generates switches whose frequency depends on  $\sigma^2$ . In theory, the autocorrelations will decline exponentially as in the linear stationary first-order AR model. In fact, using sample sizes of 2000 and 20000 it is found by simulation that the decay of the autocorrelations  $\rho_k$  is slower than exponential. Instead of a decay rate  $\rho_k = \phi^k$ ,  $|\phi| < 1$ ,  $\ln |\rho_k| / \ln k$  appears to remain constant when  $k$  increases, albeit not in every realisation. This corresponds to the theoretical form of the autocorrelations from a stationary fractionally integrated,  $I(d)$ , process. Setting  $p = \Pr\{\varepsilon_t < -1\} = \Pr\{\varepsilon_t > 1\}$ , simulations with  $T = 20000$  show how the estimated memory index  $\hat{d}$  increases from 0.251 for  $p = 0.01$  to 0.749 when  $p = 0.001$ . Clearly, from autocorrelations one could infer that the observations are generated by a linear long memory model. Interestingly, this possibility of misinterpretation is stronger in subsamples where  $T = 2000$ . There  $\hat{d} = 0.742$  already for  $p = 0.01$ . The conclusion is

that it may sometimes be difficult in practice to distinguish between nonlinear and linear models.

The authors argue that because a very simple model is already capable of producing autocorrelations of that type, the same is likely to happen with more complex nonlinear models. A good example of this is the stationary threshold autoregressive model by [52] that in many applications would constitute a relevant alternative to linear AR models with a unit root.

A question Granger and Teräsvirta do not answer is how long series are required for the decay rate of autocorrelations to be exponential, as the theory prescribes. Simulations show that a rough answer would be  $T = 250000$ .

Granger returns to this topic in [34]. The set-up is somewhat different in that the model generates breaks, but the results are quite similar to the ones in [46]. The model is again very simple:

$$y_t = m_t + \varepsilon_t \quad (29)$$

where  $\varepsilon_t$  is white noise  $(0, \sigma_\varepsilon^2)$ . Furthermore,

$$m_t = m_{t-1} + q_t \eta_t \quad (30)$$

where the mutually independent variables  $q_t$  and  $\eta_t$  are iid:  $q_t \sim \text{Bernoulli}(p)$ , and  $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$ . Combining (29) and (30) yields

$$y_T = m_0 + \sum_{t=1}^T q_t \eta_t + \varepsilon_T \quad (31)$$

where  $m_0$  is the starting-value. It is seen from (31) that the frequency of breaks must depend on  $p$  and their size on  $\sigma_\eta^2$ . The difference between (31) and (28) is that in the former model the switches are generated by the model, whereas in the latter they have exogenous causes. To make the asymptotics work, [34] assume that asymptotically there are sufficiently many breaks:  $pT \rightarrow c > 0$  as  $T \rightarrow \infty$ . This prevents the distance between adjacent breaks from becoming infinite as  $T \rightarrow \infty$ .

Simulations with  $T = 2000$  and  $\sigma_\varepsilon^2 = 1$  (Table 1 of the paper) show that for the smallest  $p = 0.0025$  (five breaks on average) and smallest variance  $\sigma_\eta^2 = 0.005$ , the estimated memory index  $\hat{d} = 0.076$ . When  $p = 0.05$  (the largest value, 100 breaks on average) and  $\sigma_\eta^2 = 0.1$ ,  $\hat{d} = 0.825$ . These results support earlier ones suggesting that in practice (linear) models with breaks on one hand and long memory on the other are often hard to distinguish from each other. Power transformations of break processes are examined in [13]. Their simulation results show that the estimated  $d$  varies little with the power, whereas  $d$  increases when the break probability  $p$  decreases.

## 7. Forecasting

### 7.1. Forecasting with nonlinear models

Forecasting has always been one of the areas of interest for Clive Granger. For an overview, see Clements, this issue. [45] and [73], Chapter 14, among others, discuss the

difficulties present in generating multi-step forecasts from nonlinear models. The forecasts considered here are made at time  $T$  and are conditional expectations given the information set  $\mathcal{F}_T$ . The minimum mean square forecast error criterion gives the optimal one-period mean forecast  $f_{T,1}^y = E\{y_{T+1}|\mathcal{F}_T\}$ . To illustrate Granger's contribution to the multi-step forecasting problem, we follow [73], Chapter 14, and consider the simple regression model

$$y_{t+1} = g(x_t) + \varepsilon_t \quad (32)$$

where  $\varepsilon_t$  is zero mean white noise. For example,

$$y_{t+1} = x_t^2 + \varepsilon_t \quad (33)$$

where  $x_t$  follows a linear AR(1)-process  $x_t = \phi x_{t-1} + \eta_t$ ,  $|\phi| < 1$ , and  $\eta_t$  has mean zero and the cumulative distribution function  $\mathcal{D}$ . Now, the forecast  $f_{T,1}^y = g(x_T)$ , as  $E\{\varepsilon_{T+1}|\mathcal{F}_T\} = 0$ . The optimum two-step forecast is

$$f_{T,2}^y = E\{y_{T+2}|\mathcal{F}_T\} = E\{g(x_{T+1})|\mathcal{F}_T\}.$$

As  $x_{T+1}$  is not usually known at time  $T$ , it has to be forecast from the AR(1)-process, which is not difficult, the result being  $f_{T,1}^x = \phi x_T$ . There are four alternative ways of computing the two-step forecast

$$f_{T,2}^y = E\{g(f_{T,1}^x + \eta_{T+1})|\mathcal{F}_T\}. \quad (34)$$

They are:

(i) naïve,

so that the presence of  $\eta_{T+1}$  in (34) is ignored by putting its value to zero. The forecast is  $fn_{T,2}^y = g(f_{T,1}^x)$ .

(ii) exact,

$$fe_{T,2}^y = \int_{-\infty}^{\infty} g(f_{T,1}^x + z)d\mathcal{D}(z).$$

For example, if  $\eta_t \sim \text{iid } \mathcal{N}(0, \sigma_\eta^2)$ , then  $\mathcal{D}(z)$  is the cumulative distribution function of the normal variable. The value of the integral is determined by numerical integration. The variance  $\sigma_\eta^2$  is usually unknown, but in practice the residual variance from the estimated model is used as an estimate. The integral can also be approximated numerically, which leads to

(iii) Monte Carlo,

$$fm_{T,2}^y = \frac{1}{N} \sum_{j=1}^N g(f_{T,1}^x + z_j)$$

where  $z_j, j = 1, \dots, N$ , are random numbers drawn independently from the distribution  $\mathcal{D}$ . Even here, estimates have to be substituted for the unknown parameters in  $\mathcal{D}$  such as the variance. For  $N$  large enough,  $fe_{T,2}^y$  and  $fm_{T,2}^y$  should be virtually identical. If  $\mathcal{D}$  is not assumed known (up to the variance), one can still apply

(iv) bootstrap,

$$fb_{T,2}^y = \frac{1}{N_B} \sum_{j=1}^{N_B} g(f_{T,1}^x + \hat{\eta}_t^{(j)})$$

where  $\hat{\eta}_t^{(j)}, j = 1, \dots, N_B$ , are the  $N_B$  independent draws with replacement from the set of residuals  $\{\hat{\eta}_t\}_{t=2}^T$  coming from the estimated AR(1) model over the sample period.

For the particular model (33), the four forecasts will be, effectively,

$$\begin{aligned} fn_{t,2}^y &= \phi^2 x_t^2 \\ fe_{t,2}^y \simeq fm_{t,2}^y &= \phi^2 x_t^2 + \sigma_\eta^2 \\ \text{and } fb_{t,2}^y &= \phi^2 x_t^2 + \hat{\sigma}_\eta^2 \end{aligned}$$

(assuming  $N_B^{-1} \sum_{j=1}^{N_B} \hat{\eta}_t^{(j)}$  is near zero). The naïve method yields biased forecasts, as  $\sigma_\eta^2$  is ignored. In practice the function  $g(\cdot)$  in (32) is not known so that its form has to be specified and the parameters estimated. Thus  $g$  has to be replaced by  $\hat{g}$  in the four forecasts above.

### 7.2. Nonlinear combination of forecasts

In a classic paper, [4] argue that a combination of forecasts often leads to a more accurate forecast than the individual ones. Since the publication of this work, various weighting schemes have been examined and their performance investigated over the years. A weighting scheme can be time-varying and the weights functions of random variables. [12] examine the situation in which the weights are nonlinear functions of them. Suppose there are two forecasts of  $y_{T+1}$ :  $f_{1,T}^y$  and  $f_{2,T}^y$ , made at time  $T$ . A nonlinear combination  $f_{C,T}^y$  of these depending nonlinearly on  $x_T$  may be written as

$$f_{C,T}^y = w(x_T)(\alpha_1 f_{1,T}^y + \alpha_2 f_{2,T}^y) + \{1 - w(x_T)\}(\alpha_3 f_{1,T}^y + \alpha_4 f_{2,T}^y) \tag{35}$$

where  $w(x_T) = (1 + \exp\{-\gamma x_T\})^{-1}$  with  $\gamma > 0$ . The weights thus change smoothly from  $\alpha_1$  and  $\alpha_2$  to  $\alpha_3$  and  $\alpha_4$  as a function of  $x_T$ . When  $\gamma \rightarrow \infty$ , the switch becomes abrupt. This alternative as well as some others are also discussed in the paper. A simple example of the weights in (35) would be the one with  $\alpha_1 = \alpha_4 = 1$  and  $\alpha_2 = \alpha_3 = 0$ .

An empirical example involving two series of UK inflation forecasts indicates that a simple linear combination does not improve forecasting accuracy compared to single forecasts whereas the nonlinear weights do. When  $\gamma < \infty$ , meaning that the transition is smooth, the slope parameter has to be selected by the user. One way of doing that is to take historical forecasts, experiment with different values of  $\gamma$  and see which one or ones work best.

## 8. Other topics

### 8.1. Chaos

The nonlinear models considered so far have been stochastic. In many branches of science, nonlinear processes in use are deterministic, such as chaos. A question then is, whether even economic time series could be regarded as generated by deterministic processes, which is the case in some economic theory models. (Granger's answer is negative, see, for example, [24]). [57] examine white chaos, a deterministic process with the same properties as linear white noise in that no linear dependence can be detected in realisations of it using normal statistical techniques. As an example, random number generators produce such series. White chaos is thus another example of a nonlinear model with misleading linear properties.

Granger and his co-authors look at a measure, the so-called correlation integral, used to detect chaos and a test based on that measure. Given a time series  $\{y_t, t = 1, \dots, T\}$ , let  $\mathbf{y}_t = (y_t, y_{t+1}, \dots, y_{t+m+1})$  be an  $m$ -vector. The correlation integral is defined using  $y_t$  as follows:

$$C_m(\varepsilon) = \lim_{T \rightarrow \infty} T^{-2} \{ \text{number of pairs } (i, j), i \neq j, \text{ such that } |y_{t+i} - y_{t+j}| < \varepsilon, \\ i, j = 0, 1, \dots, m-1 \} \quad (36)$$

For small  $\varepsilon$ , (36) grows exponentially with rate  $\nu_m$ :  $C_m(\varepsilon) \approx \exp\{\nu_m\}$ , where  $m$  is called the embedding dimension. The corresponding correlation dimension equals  $C(\varepsilon) = \exp\{\nu\}$ . For appropriate choices of  $\varepsilon$  and  $m$ ,  $C(\varepsilon)$  may be numerically approximated by  $C_m(\varepsilon)$ . The exponent  $\nu$  is constant for various models of chaos, but for stochastic white noise  $C(\varepsilon) = \exp\{m\}$ , i.e., the correlation dimension grows exponentially with the embedding dimension. Simulations of  $C_m(\varepsilon)$  in [57] support this conclusion. Applying the correlation integral to two daily financial series, the IBM and the Standard and Poor 500 index returns with 5900 observations, they find no evidence of white chaos.

The authors also examine a statistical test based on the correlation integral, the so-called BDS test by [8] (the working paper version of the test appeared in 1987). The null hypothesis of the BDS test is that the process (time series) consists of iid observations, and the alternative is that this is not true. [57] consider the size and power of the test when it is applied to the residuals of a linear autoregressive model. The data-generating process is either the logistic map with the parameter value four or any of the models considered in [53], and  $T = 200$ . It is found that the empirical size of the test varies strongly with  $\varepsilon$ : in simulations,  $\varepsilon = 0.8^j$ ,  $j = 1, 2, \dots, 10$ . Both the largest values of  $\varepsilon$  corresponding to  $j = 1, 2$  and the smallest one for which  $j = 10$  are useless in the sense that the test is strongly positively size-distorted. The results on power show that the test convincingly rejects the null hypothesis when data are generated by the logistic map. However, the test also has power against other nonlinear models. It would therefore be a mistake to claim that a rejection by the BDS test would be an indication of white chaos.

[57] remark that if one wants to distinguish between white chaos and stochastic white noise, one would need a statistical test in which chaos is the null hypothesis. Unfortunately,

such a test does not exist. [25], discussing the paper by [9], points out that the difference between a stochastic iid (assuming white noise is not enough) process and white chaos becomes clear in forecasting. While the former is never forecastable, the latter can be forecast perfectly if the starting-values are known and there is no measurement error.

## 8.2. Aggregation of nonlinear series

The set-up in [53] is used again by [40] who are studying the effects of aggregation on nonlinearity. They construct a situation in which there is nonlinearity at the micro level, but only either the cross-sectionally or temporally aggregated series is actually observed. In cross-sectional aggregation the micro equation has the form

$$y_{jt} = f(y_{j,t-1}, y_{j,t-2}, \dots; \varepsilon_{j,t-1}, \varepsilon_{j,t-2}, \dots) + \varepsilon_{jt}$$

where the zero mean white noise error process  $\varepsilon_t = e_t + \eta_{jt}$ , with the common error  $e_t$  and an idiosyncratic one,  $\eta_{jt}$ , where  $\eta_{jt}$  and  $\eta_{kt}$ ,  $j \neq k$ , are independent. All have finite variances. The bilinear model (7), i.e.,

$$y_{jt} = \theta_{12} y_{j,t-2} \varepsilon_{j,t-1} + \varepsilon_{jt}$$

serves as a convenient theoretical example as it is linear in parameters. A cross-sectional aggregation of  $N$  series leads to

$$Y_t = \theta_{12} e_{t-1} Y_{t-2} + N e_t + \theta_{12} \sum_{j=1}^N y_{j,t-2} \eta_{j,t-1} + \sum_{j=1}^N \eta_{jt} \quad (37)$$

where  $Y_t = \sum_{j=1}^N y_{jt}$ . Note that the parameter  $\theta_{12}$  is the same across the equations. Effects of aggregation are controlled by the balance between the components  $N e_t$  and the sum  $\theta_{12} \sum_{j=1}^N y_{j,t-2} \eta_{j,t-1} + \sum_{j=1}^N \eta_{jt}$ . The former one has variance of order  $O(N^2)$ , whereas the latter ones, due to the independence of  $\eta_{jt}$  and  $\eta_{kt}$ , has variance of order  $O(N)$ . Consequently, for large  $N$  the term  $N e_t$  will dominate, and the model (37) for  $Y_t$  is still approximately bilinear, as the last two sums in (37) can be ignored. If  $e_t = 0$ , meaning that there is no common component in the errors, the two sums play a bigger role, and the model can no longer be recognised as bilinear.

The authors conduct a simulation study using the tests and models in [53] and varying the variance of  $e_t$  and  $\eta_{jt}$  (the latter is assumed to be identical for all  $\eta_{jt}$ ). The results are in line with the ones in the aforementioned example and support the claim that aggregation weakens nonlinearity present on the micro level. Temporal aggregation is also considered, and the results are similar.

## 8.3. Nonlinear dependence and trends

It is not difficult to find situations in which linear dependence measures indicate no dependence, but the variables in question are nonlinearly dependent. Using standard

autocorrelation coefficients to determine the maximum lag in a potentially nonlinear univariate model may in those cases be misleading. [41] study lag selection in nonlinear moving average and autoregressive models. The authors apply the mutual information coefficient which is a function of Shannon entropy normalised to lie between zero and one. Computing this dependence measure requires nonparametric density estimation, which is discussed in the paper. Simulations show the usefulness of this measure in determining the lag structure of these models.

[42] look at another measure of nonlinear dependence based on the normalised Bhattacharyya–Matusita–Hellinger dependence measure. They demonstrate its connection with yet another dependence measure: the copula. This measure is also computed using kernel estimation. These authors examine its performance using the same set of models as [41] and find it fully satisfactory. In addition they consider chaos in the form of the logistic map. They find that linear autocorrelation estimates do not indicate any dependence in the series generated by the logistic map, which agrees with the results in [57]. On the contrary, their dependence measure provides strong evidence against independence. More discussion on nonlinear dependence can be found in [73], Chapter 4.

[1] were the first econometricians to examine common nonlinear features in multiple equation systems. [47] extend the definition of common features to distributions and apply copulas to investigate the nonlinear dependence of consumption and income on the business cycle. Their results 'give some support to the claim that the impact of the business cycle on the joint distribution of consumption and income is through the marginal distributions and not through their dependence structure'. This makes the business cycle a 'common factor in distribution' for consumption and income.

Finally, it should be mentioned that Granger has always been very interested in trends. In [23] he considers models that generate linear or nonlinear trends. In [36] the authors look at both stochastic and deterministic nonlinear trends related to growth processes. The paper [79], written jointly with Hal White, contains a wealth of information on trends: history, definitions, generating mechanisms, estimation, forecasting and so on. Anyone interested in trends cannot afford to miss it.

## 9. Final remarks

This paper allows a glimpse on the large amount of work Clive Granger has done in the area of nonlinear models and modelling. An attempt is made to describe a few of the new ideas related to nonlinearity that Granger has produced over the years. Many of them can be pursued further, and it is hoped that this will be the case in the future.

## Acknowledgements

Work for this paper has been supported by the Center for Research in Econometric Analysis of Time Series (CREATES), funded by the Danish National Research Foundation, Grant No. DNFR 78. I wish to thank two anonymous referees for useful remarks. Responsibility for any errors or shortcomings in this work remains mine.

### References

- [1] Heather M. Anderson and Farshid Vahid. Testing multiple equation systems for common nonlinear components. *Journal of Econometrics*, 84:1–36, 1998.
- [2] R. Ashley, C. W. J. Granger, and R. Schmalensee. Advertising and aggregate consumption: An analysis of causality. *Econometrica*, 48:1149–1167, 1980.
- [3] Nathan S. Balke and Thomas B. Fomby. Threshold cointegration. *International Economic Review*, 38:627–645, 1997.
- [4] J. M. Bates and Clive W. J. Granger. The combination of forecasts. *Operational Research Quarterly*, 20:451–468, 1969.
- [5] Ralf Becker and A. Stan Hurn. Testing for nonlinearity in mean in the presence of heteroskedasticity. *Economic Analysis & Policy*, 39:311–326, 2009.
- [6] Anil K. Bera and Matthew L. Higgins. ARCH and bilinearity as competing models for nonlinear dependence. *Journal of Business & Economic Statistics*, 15:43–50, 1997.
- [7] George E. P. Box and Gwilym M. Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, 1970.
- [8] W. A. Brock, W. D. Dechert, J. A. Scheinkman, and B. LeBaron. A test for independence based on the correlation dimension. *Econometric Reviews*, 15:197–235, 1996.
- [9] Sangit Chatterjee and Mustafa R. Yilmaz. Chaos, fractals and statistics. *Statistical Science*, 7:49–68, 1992.
- [10] Harald Cramér. On some classes of non-stationary processes. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pages 157–178. University of California Press, 1961.
- [11] R. B. Davies. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64:247–254, 1977.
- [12] Melinda Deutsch, Clive W. J. Granger, and Timo Teräsvirta. The combination of forecasts using changing weights. *International Journal of Forecasting*, 10:47–57, 1994.
- [13] Ingolf Dittmann and Clive W. J. Granger. Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics*, 110:113–133, 2002.
- [14] Bruno Eklund. *Four contributions to statistical inference in econometrics*. Stockholm School of Economics, Stockholm, 2003.

- [15] Walter Enders and Clive W. J. Granger. Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. *Journal of Business & Economic Statistics*, 16:304–331, 1998.
- [16] Luigi Ermini and Clive W. J. Granger. Some generalizations on the algebra of I(1) processes. *Journal of Econometrics*, 58:369–384, 1993.
- [17] Eric Ghysels, Norman R. Swanson, and Mark W. Watson. Introduction. In Eric Ghysels, Norman R. Swanson, and Mark W. Watson, editors, *Essays in econometrics. Collected papers of Clive W.J. Granger*, volume 1, pages 1–27, Cambridge, 2001. Cambridge University Press.
- [18] Stephen M. Goldfeld and Richard E. Quandt. *Nonlinear Methods in Econometrics*. North-Holland, Amsterdam, 1972.
- [19] Clive W. J. Granger. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14:227–238, 1980.
- [20] Clive W. J. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16:121–130, 1981.
- [21] Clive W. J. Granger. Forecasting white noise. In Arnold Zellner, editor, *Applied time series analysis of economic data : Proceedings of the Conference on Applied Time Series Analysis of Economic Data, October 13-15, 1981, Arlington, VA*, pages 308–314, Washington, DC, 1983. U.S. Department of Commerce, Bureau of the Census.
- [22] Clive W. J. Granger. Developments in the study of cointegrated variables. *Oxford Bulletin of Economics and Statistics*, 48:213–228, 1986.
- [23] Clive W. J. Granger. Models that generate trends. *Journal of Time Series Analysis*, 9:329–343, 1988.
- [24] Clive W. J. Granger. Developments in the nonlinear analysis of economic series. *Scandinavian Journal of Economics*, 93:263–276, 1991.
- [25] Clive W. J. Granger. Comment. *Statistical Science*, 7:102–104, 1992.
- [26] Clive W. J. Granger. Strategies for modelling nonlinear time-series relationships. *Economic Record*, 69:233–238, 1993.
- [27] Clive W. J. Granger. Some comments on empirical investigations involving cointegration. *Econometric Reviews*, 13:345–350, 1994.
- [28] Clive W. J. Granger. Modelling nonlinear relationships between extended-memory variables. *Econometrica*, 63:265–279, 1995.

- [29] Clive W. J. Granger. Overview of nonlinear macroeconomic models. *Macroeconomic Dynamics*, 5:466–481, 2001.
- [30] Clive W. J. Granger. Non-linear models: Where do we go next - time-varying parameter models? *Studies in Nonlinear Dynamics and Econometrics*, 12:Issue 1, Article 3, 2008.
- [31] Clive W. J. Granger and A. P. Andersen. *An Introduction to Bilinear Time Series*. Vandenhoeck and Ruprecht, Göttingen, 1978.
- [32] Clive W. J. Granger and Allan P. Andersen. On the invertibility of time series models. *Stochastic Processes and their Applications*, 8:87–92, 1978.
- [33] Clive W. J. Granger and Jeff Hallman. Nonlinear transformations of integrated time series. *Journal of Time Series Analysis*, 12:207–224, 1991.
- [34] Clive W. J. Granger and Namwon Hyung. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11:399–421, 2004.
- [35] Clive W. J. Granger and Namwon Hyung. Introduction to m-m processes. *Journal of Econometrics*, 130:143–164, 2006.
- [36] Clive W. J. Granger, Tomoo Inoue, and Norman Morin. Nonlinear stochastic trends. *Journal of Econometrics*, 81:65–92, 1997.
- [37] Clive W. J. Granger, Maxwell L. King, and Halbert White. Comments on testing economic theories and the use of model selection criteria. *Journal of Econometrics*, 67:173–187, 1995.
- [38] Clive W. J. Granger and Hahn S. Lee. An introduction to time-varying parameter cointegration. In Peter Hackl and Anders H. Westlund, editors, *Economic structural change. Analysis and forecasting*, pages 139–157, Berlin, 1991. Springer.
- [39] Clive W. J. Granger and T. H. Lee. Investigation of production, sales and inventory relationships using multicointegration and non-symmetric error using correction models. *Journal of Applied Econometrics*, 4:S145–S159, 1989.
- [40] Clive W. J. Granger and Tae-Hwy Lee. The effect of aggregation on nonlinearity. *Econometric Reviews*, 18:259–269, 1999.
- [41] Clive W. J. Granger and Jin-Lung Lin. Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis*, 15:371–384, 1994.
- [42] Clive W. J. Granger, Esfandiar Maasoumi, and Jeff Racine. A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis*, 25:649–669, 2004.

- [43] Clive W. J. Granger and Norman Swanson. Further developments in the study of cointegrated variables. *Oxford Bulletin of Economics and Statistics*, 58:537–553, 1996.
- [44] Clive W. J. Granger and Norman R. Swanson. An introduction to stochastic unit-root processes. *Journal of Econometrics*, 80:35–62, 1997.
- [45] Clive W. J. Granger and Timo Teräsvirta. *Modelling Nonlinear Economic Relationships*. Oxford University Press, Oxford, 1993.
- [46] Clive W. J. Granger and Timo Teräsvirta. A simple nonlinear time series model with misleading linear properties. *Economics Letters*, 62:161–165, 1999.
- [47] Clive W. J. Granger, Timo Teräsvirta, and Andrew J. Patton. Common factors in conditional distributions for bivariate time series. *Journal of Econometrics*, 132:45–57, 2006.
- [48] David Harris, Brendan McCabe, and Stephen Leybourne. Stochastic cointegration: estimation and inference. *Journal of Econometrics*, 111:363–384, 2002.
- [49] David F. Hendry. The Nobel Memorial Prize for Clive W.J. Granger. *Scandinavian Journal of Economics*, 106:187–213, 2004.
- [50] Anders B. Kock and Timo Teräsvirta. Forecasting performance of three automated modelling techniques during the economic crisis 2007-2009. *International Journal of Forecasting*, 30:616–631, 2014.
- [51] Anders B. Kock and Timo Teräsvirta. Forecasting macroeconomic variables using neural network models and three automated model selection techniques. *Econometric Reviews*, 35:1753–1779, 2016.
- [52] Markku Lanne and Pentti Saikkonen. Threshold autoregressions for strongly autocorrelated time series. *Journal of Business & Economic Statistics*, 20:282–289, 2002.
- [53] Tae-Hwy Lee, Halbert White, and Clive W. J. Granger. Testing for neglected non-linearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics*, 56:269–290, 1993.
- [54] S. J. Leybourne, B. P. M. McCabe, and A. R. Tremayne. Can economic time series be differenced to stationarity? *Journal of Business & Economic Statistics*, 14:435–446, 1996.
- [55] Stephen J. Leybourne, Brendan P. M. McCabe, and Terence C. Mills. Randomized unit root processes for modelling and forecasting financial time series: Theory and applications. *Journal of Forecasting*, 15:253–270, 1996.

- [56] Chien-Fu J. Lin and Timo Teräsvirta. Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics*, 62:211–228, 1994.
- [57] T. Liu, C. W. J. Granger, and W. P. Heller. Using the correlation exponent to decide whether an economic series is chaotic. *Journal of Applied Econometrics*, 7:S25–S39, 1992.
- [58] Ritva Luukkonen, Pentti Saikkonen, and Timo Teräsvirta. Testing linearity against smooth transition autoregressive models. *Biometrika*, 75:491–499, 1988.
- [59] Agustin Maravall. An application of nonlinear time series forecasting. *Journal of Business & Economic Statistics*, 1:66–74, 1983.
- [60] Brendan McCabe, Stephen J. Leybourne, and David Harris. A residual-based test for stochastic cointegration. *Econometric Theory*, 22:429–456, 2006.
- [61] Brendan P. M. McCabe and Andrew R. Tremayne. Testing a time series for difference stationarity. *Annals of Statistics*, 23:1015–1028, 1995.
- [62] A. I. McLeod and W. K. Li. Diagnostic checking ARMA time series models using squared residual autocorrelations. *Journal of Time Series Analysis*, 4:269–273, 1983.
- [63] Maurice B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society, Series B*, 27:204–237, 1965.
- [64] B. G. Quinn. Stationarity and invertibility of simple bilinear models. *Stochastic Processes and their Applications*, 12:225–230, 1982.
- [65] Antti Ripatti and Pentti Saikkonen. Vector autoregressive processes with nonlinear time trends in cointegrating relations. *Macroeconomic Dynamics*, 5:577–597, 2001.
- [66] Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.
- [67] Gideon Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 4:461–464, 1978.
- [68] James H. Stock and Mark W. Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In Robert F. Engle and Halbert White, editors, *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, pages 1–44, Oxford, 1999. Oxford University Press.
- [69] T. Subba Rao and M. M. Gabr. *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Springer, New York, 1984.
- [70] Timo Teräsvirta. Power properties of linearity tests for time series. *Studies in Nonlinear Dynamics and Econometrics*, 1:3–10, 1996.

- [71] Timo Teräsvirta, Chien-Fu Lin, and Clive W. J. Granger. Power of the neural network linearity test. *Journal of Time Series Analysis*, 14:309–323, 1993.
- [72] Timo Teräsvirta, Dag Tjøstheim, and Clive W. J. Granger. Aspects of modelling nonlinear time series. In Robert F. Engle and Daniel L. McFadden, editors, *Handbook of Econometrics*, volume 4, pages 2919–2957, Amsterdam, 1994. Elsevier.
- [73] Timo Teräsvirta, Dag Tjøstheim, and Clive W. J. Granger. *Modelling Nonlinear Economic Time Series*. Oxford University Press, Oxford, 2010.
- [74] Howell Tong. On a threshold model. In Chi-Hau Chen, editor, *Pattern recognition and signal processing*, number 29 in Nato Advanced Study Institutes series: Applied sciences. Sijthoff & Noordhoff, 1978.
- [75] Howell Tong. *Non-Linear Time Series. A Dynamical System Approach*. Oxford University Press, Oxford, 1990.
- [76] Ruey S. Tsay. Nonlinearity tests for time series. *Biometrika*, 73:461–466, 1986.
- [77] Halbert White. An additional hidden unit test for neglected non-linearity in multi-layer feedforward networks. In *Proceedings of the International Joint Conference on Neural Networks, Washington DC*, volume 1, pages 451–55, San Diego, CA, 1989. SOS Printing.
- [78] Halbert White. Approximate nonlinear forecasting methods. In Graham Elliott, Clive W. J. Granger, and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 1, pages 459–512, Amsterdam, 2006. Elsevier.
- [79] Halbert White and Clive W. J. Granger. Consideration of trends in time series. *Journal of Time Series Econometrics*, 3:Iss. 1, Article 2, 2011.
- [80] Gawon Yoon. A note on some properties of STUR processes. *Oxford Bulletin of Economics and Statistics*, 68:253–260, 2006.
- [81] Gawon Yoon. Stochastic unit root processes: Maximum likelihood estimation, and new Lagrange multiplier and likelihood ratio tests. *Economic Modelling*, 52:725–732, 2016.