MONASH

UNIVERSITY

AUSTRALIA

STRATEGIES FOR MODELLING NONLINEAR TIME SERIES RELATIONSHIPS

Clive W.J. Granger

Working Paper No. 1/93

January 1993

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0369 2

STRATEGIES FOR MODELLING NONLINEAR TIME SERIES RELATIONSHIPS

Clive W.J. Granger

Working Paper No. 1/93

January 1993

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS COMMERCE & MANAGEMENT

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

STRATEGIES FOR MODELLING NONLINEAR TIME SERIES RELATIONSHIPS[1]

Clive W.J. Granger

Preamble

This paper was presented to the Australian meeting of the Econometries Society at Monash University in July 1992 as the A.W. Phillips lecture. It is dedicated to the memory of Bill Phillips and to the excellent, fundamental work that he produced in various parts of economics. Whilst investigating the relationship between unemployment and wage rates, generally known as the Phillip's curve, he considered several alternative non–linear specifications although he was limited by having very small samples and very little computing power. Present applied workers are less limited and, when considering non–linear modelling have a variety of specifications that can be considered. In this paper, I discuss the kind of problems that they face and try to suggest strategies that can be used. Modelling questions in a non–linear framework are quite likely to be very different from those face when building linear models. The paper is thus rather unconventional as it attempts to present opinions and to initiate discussion of this potentially important area rather than to present specific results or theorems.

[1] Supported by NSF grant SES 9023037.

1. Introduction

It is often not an easy task to build a linear model, as the modeller faces questions about what variables and what lags to use, but there are many more problems when nonlinear modelling of a relationship is attempted. This became clear to me whilst writing a book on the area (Granger and Teräsvirta (1993)) and a related chapter for the Handbook of Econometrics (Teräsvirta, Tjostheim and Granger (1992)). Consider a multiple input, single output situation

$$Y_{t+1} = F(\underline{W}_t) + \varepsilon_{t+1} \qquad (1)$$

where the vector $\underline{W}_t$ has M components and consists of lags of Y_t and lagged components of a vector $\underline{X}_t$ of m explanatory variables. If lags up to order p are allowed then $M = (m+1)p$. This is the number of parameters that would be fitted in a linear model (excluding the variance of ε_t) and, if large, gives the usual "curse of dimension." It is convenient to say that the relationship is <u>linear in mean</u> (with respect to $\underline{W}_t$) if

$$E[Y_{t+1} \mid \underline{W}_t] = a' \underline{W}_t$$

so that no nonlinear terms occur in this conditional mean. In this paper, only a single dependent variable will be considered, rather than a full vector or system and only reduced form equations, as in (1). If contemporaneous values of the explanatory values were included, such as $\underline{X}_{t+1}$, one would have a "structural" equation with all the usual problems of interpretation and identification, although these are more complicated in the non–linear case. For reasons to be explained below, evaluation of such models are particularly difficult. The objective of the modelling exercise will be taken to be to achieve an approximation to the true generating mechanism for the conditional mean, using $\underline{W}_t$ as the information set. The model will automatically provide useful one–step forecasts but forecasts over longer horizons will require further modelling.

The question considered here is how should an applied worker proceed to try to build a model such as (1). The problem is likely to arise as economists, particularly theorists, appear to believe that relationships between economic variables are interestingly nonlinear. If a theory completely specifies a relationship, such as the function in (1), apart from some unknown parameters $\underline{\theta}$, then the econometrician can apply a maximum likelihood procedure to estimate $\underline{\theta}$ and some large–sample properties of the estimate are known. However, it is often the case that the theory does not completely specify the non–linear relations, production functions being an example, so that the applied worker is left with the task of finding the function F in (1), or at least an acceptable approximation for it. The task is a daunting one — as there are obviously a huge number of possible functions to consider. Even if some natural constraints, such as for stability, are imposed the number of possible functions remains immerse.

It will be assumed that data is available on Y_{t+1}, $\underline{W}_t$ for $t = 1, \dots, N$.

A useful example which can be used to illustrate the kind of problem faced by applied workers is a model which explains electricity demand for some region in terms of temperature and other variables. It is well known that the demand for electricity increases as temperatures become high, due to the use of air–conditioning, and that demand also increases as temperatures fall to low levels because of the use of heating appliances. There is thus a substantial nonlinear relationship between temperature and electricity demand. The general features of this relationship is understood but no specific functional form comes from the use of economic theory, as far as I am aware. A study of this relationship is found in Engle et al. (1986).

2. Non–linear Modelling

Suppose now that the applied worker decides to consider fitting a non–linear model to the data, with the specification not given completely by a theory. At this point there are two basic strategies:

S1: test a null hypothesis of linearity, and if this is rejected, go to S2, and

S2: chose a particular non–linear model, or some small group of models, estimate, analyze and evaluate them. Use a model selection criterion to chose the "best" model from the group.

In this section just the second of these strategies is considered and in section 3 the discussion returns to S1, which is recommended.

The immediate problem facing the researcher is that there are many possible nonlinear models. These include:

Bilinear

Y_{t+1} modelled in terms of products, such as $Y_{t-j}\varepsilon_{t-k}$, $X_{i,t-j}\varepsilon_{t-k}$ etc. plus linear terms. (2)

Flexible Fourier Form

$$Y_{t+1} = \underline{\alpha}'\underline{W}_t + \sum_{j=1}^{q} \gamma_j \operatorname{cosine}(\underline{\beta}_j'\underline{W}_t + \theta_j) + e_{t+1} \tag{3}$$

Neural Network

$$Y_{t+1} = \underline{\alpha}'\underline{W}_t + \sum_{j=1}^{q} \gamma_j\, \phi(\underline{\beta}_j'\underline{W}_t + \theta_j) + e_{t+1} \tag{4}$$

where $\phi(z)$ is a "squashing function," usually a bounded, monotonic function such as the logistic function $\phi(z) = (1 + \exp(-z))^{-1}$, so that $\phi(-\infty) = 0$, $\phi(\infty) = 1$.

Projection Pursuit

$$Y_{t+1} = \underline{\alpha}'\underline{W}_t + \sum_{j=1}^{q} \gamma_j\, \phi_j(\underline{\beta}_j'\underline{W}_t + \theta_j) + e_{t+1} \tag{5}$$

where $\phi_j(z)$ is estimated nonparametrically for given z, using a "smoother" of some form. In fact, a sequence of models are fitted with q increasing, the model at one stage being subtracted from Y_{t+1} and another non–linear term then fitted to the residual. A stopping rule is then applied to find q.

Smooth Regime Switching (or Smooth Transition Regression)

$$Y_{t+1} = \underline{\alpha}_1'\underline{W}_t + \phi(\underline{\gamma}'\underline{W}_t)\, [\underline{\alpha}_2'\underline{W}_t] \tag{6}$$

where $\phi(z)$ is bounded in (o, 1) and monotonic, such as the logistic function defined above. In this case the model switches smoothly from one linear regime R_1:$Y_{t+1} = \underline{\alpha}_1'\underline{W}_t + e$ to another regime R_2:$Y_{t+1} = (\underline{\alpha}_1'+\underline{\alpha}_2')\underline{W}_t + e$. $\underline{\gamma}'\underline{W}_t$ is the "switching variable" and may contain just a single term and can then approximate sharp switching models such as STAR. The regimes considered here are linear, but need not be. The model has the advantage that it can be associated with various parts of economic theory, which suggests changing relationships as industrial spare capacity becomes low or unemployment becomes low or as the economy swings from a peak to a trough. All of these models, and others, are described in detail in Granger and Teräsvirta (1993).

Time Varying Parameter (TVP) Models

$$Y_{t+1} = \underline{\alpha}_t \underline{W}_t + e_{t+1} \qquad (7)$$

where $\underline{\alpha}_t$ varies stochastically but not specifically as a function of $\underline{W}_{t-j}$, such as

$$\underline{\alpha}_t = \underline{D}\ \underline{\alpha}_{t-1} + \underline{\eta}_t \qquad (8)$$

where $\underline{\eta}_t$ is a vector white noise. (8) may include unit roots. The time–varying parameters may be estimated using the Kalman algorithm. The TVP model is not specifically non–linear although it is effectively so through the working of the algorithm. It is clear that this model may well provide a good approximation to several of the specific non–linear models mentioned above.

This listing by no means exhausts the possible non–linear models that might be considered.

A number of specification problems arise, how to decide what variables go into X, how large should be p (the number of lags), how large should be q (number of non–linear terms), which model to use and how to distinguish between a specific non–linear model and the TVP model. In a neural network model there are (excluding σ_ε^2)

$$M = p(m + 1)\ (q + 1) \text{ parameters.}$$

A possibly relevant model selection criterion which can help decide the size of p and q is the Rissanen complexity criterion (Rissanen (1987))

$$\log \sigma_\varepsilon^2(p,q) + M \frac{\log\ n}{n} \qquad (9)$$

where n is the sample size and $\sigma_\varepsilon^2(p,q)$ is the variance of the residuals in (4) when particular values of p and q are used. The criterion penalizes the models which are over–parameterized and so gives some value to parsimony. As an example, if $p = 3$, $m = 3$, $q = 4$ then $M = 49$ parameters. This high value for a fairly modest model size may be thought of as the "curse of complexity." If parsimony is considered to be really important then perhaps a "super–parsimonious" criterion may be helpful, such as (9) but M replaced by M^d, for some $d > 1$, so that the criterion is

$$\log \sigma_\varepsilon^2(p,q) + M^d \frac{\log n}{n}. \tag{10}$$

Some experimentation is required before an appropriate value of d can be recommended. Clearly other more parsimonious criteria could be considered.

The number of possible non–linear generating mechanisms is immense, possibly even infinite. A good non–linear model would be able to approximate them all – which is rather unrealistic. For some of the models any function F in (1) can be well approximated for large enough q, an example being the neural network. Clearly good approximating non–linear models have to be highly flexible and so are inclined to pick up any subtle nuance in the data and if the size q is not held in close control, these models will be inclined to overfit in sample. A small simulation experiment (Granger and Teräsvirta (1992b)), using neural networks and projection pursuit with $m = 0$ or 1, $q = 2$ or 3 and $p = 1$ or 2, found that if the true generating mechanism was linear (in mean) with true variance of residual $\sigma_\varepsilon^2 = 1$, the fitted models usually over–fitted and thus found $\hat{\sigma}_\varepsilon < 1$. These models then forecast very poorly. However, for simple non–linear generating mechanisms the non–linear models both fitted well in sample and forecast well out of sample. The possibility of these non–linear models overfitting in sample suggests the following strategy:

S3: any model should be evaluated in terms of its out of sample forecasting ability, and compared with the forecasts from linear and other non–linear models. There are standard, well understood tests for comparing forecasts.

I would strongly recommend that this strategy is always followed. There are, however, problems with it. There is the usual difficulty that the parameters of the generating mechanism may have changed from the in sample to the out of sample periods, so that the test is inappropriate. Even if this does not occur, it is true that quite a long post sample period may be required for the extra forecasting ability of any particular non–linear model to become evident. For example, a regime–switching model may not forecast any better than a linear model if the switching variable stays nearly constant in the post sample period. We have experience with this situation where a regime switching model fitted well in sample but did not forecast better than a linear model out of sample, but where no switching occurred. It is difficult to propose a specific rule but I certainly think that, say, 20% of any sample should be held back for a post–sample, forecasting evaluation. Strategy S3 is certainly controversial, as many researchers would prefer to keep all data in sample and to rely on various specification tests, constancy of recursive estimates of parameters and other model evaluation techniques. As so often with such a new area of research, further experience is required. It may be possible to use cross–validation, but the usual problems arise when this technique is used with time series data.

Virtually all of the non–linear models considered in this section are designed for use with stationary variables. Research is needed on the effects of I(1) variables.

3. <u>Testing for Linearity</u>

A standard linearity test takes the form of the regression

$$Y_{t+1} = \underline{\alpha}' \underline{W}_t + \text{non–linear terms} + e_{t+1}$$

with a test of significance, such as a Lagrange–multiplier test being applied to the non–linear terms. The non–linear terms can be powers of the components of $\underline{W}_t$ (Teräsvirta), or powers of the best linear models for y_{t+1} (RESET test) or a sum of p squashing functions with coefficients chosen at random (White's neural network test). These and other tests are discussed in Lee et al. (1993), who also explore the power of various tests by a Monte Carlo study. Some of the non–standard tests are proposed by Brock, Deckert and Scheinkman (the BDS test) which originated from chaos theory and one based on the bi–spectrum. Lee et al. (1993) found that no one test dominated the others, that a few tests, including the neural network test, had good power but even the best tests had poor power against certain types of non–linearity. Teräsvirta (1990) finds that a simple test, just using quadratic and cubic terms of the components of $\underline{W}_t$, usually has a good power as the more complied test procedures. Various aspects of testing are discussed in (Granger and Teräsvirta (1993)).

A clear problem with most of these tests are that they can easily be confused by heteroskedasticity, particularly of the ARCH form. This is certainly true of the non–standard tests, which are looking for any type non–linearity, not just non–linearity in mean. Sin and White (1992) have proposed an m–type (Wooldridge) test for linearity that is designed to be robust against heteroskedasticity. A linear model is fitted with ARCH residuals, $Y_{t+1} = \underline{\alpha}' \underline{W}_t + \varepsilon_t$ with var $\varepsilon_t \simeq h_t = a_0 + a_1 \varepsilon_{t-1}^2 + a_2 h_{t-1}$, say. Let M_{1t}, M_{2t} be a pair of non–linear functions of W_t, such as $(\hat{\underline{\alpha}}' \underline{w}_t)^2$, $(\hat{\underline{\alpha}}' \underline{w}_t)^3$ corresponding to a RESET test or a pair of neural network terms, and define $M_{jt}^* = M_{jt}/(h_t)^{\frac{1}{2}}$ j = 1, 2, $Y_t^* = Y_t/(h_t)^{\frac{1}{2}}$, $\varepsilon_t^* = \varepsilon_t/(h_t)^{\frac{1}{2}}$ and $c_t^* = (h_t)^{\frac{1}{2}}$. The test regresses ε_t^* on M_{jt}^*, j = 1, 2, c_t^*

and Y_t^*. Under the null of linearity, nR^2 is chi–squared with two degrees of freedom. A simulation found that when the neural test was used, it generally had the proper size, had good power against some forms of non–linearity, but not all forms.

Because of problems with heteroskedasticity, which is fairly common with economic data, it follows that non–standard tests, such as the BDS and bispectrum tests, which can be badly confused by heteroskedasticity, are of limited use if one is interested in testing for linearity in mean.

It might be suggested that a consequence of the lack of power of most tests against some types of non–linearity is that it is not appropriate to rely on a single test but that a battery of different tests needs to be utilized. For example, some of the standard tests, such as that based on neural networks, have low power against bilinear models, but an LM test can be devised to have good power against bilinear models. Thus using both tests widens the class of non–linear generating mechanisms that can be detected. Clearly, a great deal more research is required to find out what combinations of tests are superior and what types of non–linearity remain undetected, or are detected with unsatisfactory power.

Testing for linearity is obviously an extremely difficult situation, the null hypothesis of linearity involves very large number of parameters, the alternative involves a huge number of models, each with many parameters. This is clearly not a simple, standard testing situation. Further, these testing problems often have nuisance parameters defined only under the alternative hypothesis.

4. Beliefs and a Proposed Strategy

As the process of understanding how to model non–linear relationships is in an early stage it is necessary to have a starting strategy. As experience accumulates a better strategy should evolve. I think that it is useful to base the starting strategy on some simplistic beliefs, based on what experience is so far available. It will be assumed that the series involved are stationary.

Useful Simplifying Beliefs:

In the general model

$$Y_{t+1} = F(\underline{W}_t) + \varepsilon_{t+1}$$

where ε_{t+1} is white noise orthogonal in mean with $\underline{W}_t$ but possibly with heteroskedasticity, the function F( ) is smooth throughout the appropriate range. There are thus no obvious bifurcations or catastrophe occurrences. If this belief is correct there are two important consequences:

a) A simple non–linear model (that is, one with a small q value) will provide an adequate approximation to the actual generating mechanism. There is thus a rather small gain in moving from a very simple model to a much more complex one. This may suggest that a super–parsimonious criterion, such as (10), should be used, with d = 1.5 or 2, say.

b) Several non–linear models will provide almost identical approximations to the true generating mechanism. From their forms it seems that the Fourier, neural networks and projection pursuit are all likely to be similar. It follows that there is little or no reason to use all of these models, and so it may be a good strategy to concentrate on just the neural network models, say.

Given these statements and the experience that has so far been accumulated, I would like to suggest the following strategy for an applied worker contemplating building a non–linear model of the relationship between two or more economic variables using a mid–sized sample of say N = 200 to 300:

1) Start with a small set $\underline{X}_t$ of explanatory variables, say one or two, selected as being the most likely to be relevant, unless some very convincing economic theory suggests that more variables should be used.

2) Perform two or more tests of linearity, preferably tests that are robust against heteroskedasticity. These tests need not be interpreted formally, via Bonferroni probability inequalities, but rather as indications of whether or not a non–linear model is likely to prove superior to a linear model.

3) *If* the tests suggest the presence of non–linearity (in mean) then chose one or two modelling techniques and fit simple, parsimonious forms of these models. My own choice would be a smooth regime–switching regression (because of ease of interpretation) and either a neural network or a projection pursuit model, depending on the availability of a computer program, as these are models which I think are likely to be successful in finding non–linearity if it is present. I would consider using a super–parsimonious model selection criterion, in order to reduce the chance of serious data–mining.

4) When performing the tests and estimating models I would prefer to hold back a substantial amount of "post–sample" data, say 20% of the sample. The one–step forecasting ability of the non–linear models and of a (constant parameter) parsimonious linear model should be compared over the

post–sample period. A test of significant difference of the variances of the forecast errors should be applied. It would be interesting to also compare the multi–step forecasts of the models, but this is not simple or straight forward with non–linear models.

5) I would test the parameters of the linear model for constancy. If this was rejected a simple time–varying parameter model such be estimated using the Kalman algorithm. Comparison should be made of the out–of–sample one–step forecasting ability of this model with the other models. The parameters would be continually updated as each new piece of data becomes available.

6) Accept any apparent non–linearity in the relationship only if the evidence in its favour is clear–cut, particularly when using aggregate macro–data. Theory suggests that cross–sectional and also temporal aggregation is likely to reduce any actual non–linearity in the relationship.

7) If the sample size is large, say N = 500, start with such a simple analysis and then consider more complicated models.

This strategy is most appropriate, in my opinion, with stationary (I(0)) series. If series appear to be I(1) or long memory it becomes appropriate to switch to changes and error–correction terms from linear or non–linear cointegrations for the modelling process. A great deal more research is required in this area, particularly the use of error–correction terms from non–linear cointegrations, as discussed in Granger and Hallman (1991).

I am sure that it is possible to suggest non–linear models for which the above strategy would not be successful. If such models are considered to be relevant for economics then the strategy should be changed. This doubt does not apply to simple, low–dimensional attractor chaos models which have been shown to be very well approximated by neural network models.

There are many research problems outstanding, including –

i) The construction of tests robust against heteroskedasticity, knowledge of power against various types of non–linearity, how best to combine such tests;

ii) Knowledge of the relative ability of different models to approximate various non–linear generating mechanisms – for which they do perform well and for which they do not;

iii) An accumulation of practical experience on how well the various models perform with actual economic data. Are there many examples where non–linearity seems to be present. There certainly are good examples, including relating a regional electricity demand to temperature, and many suggestive examples, particularly from the stock market;

iv) How long–memory processes can be best handled in this framework.

Bibliography

Engle, R.F., C.W.J. Granger, J. Rice and A. Weiss (1986). "Semi–parametric estimates of the relation between weather and electricity sales." *Journal of American Statistical Association* 81, 310–320.

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

Granger, C.W.J. and T. Teräsvirta (1992). "Experiments in modeling nonlinear relationships between time series," in M. Casdagli and S. Eubank (eds)., *Nonlinear modeling and forecasting.* Addison–Wesley, Redwood City, CA.

Granger, C.W.J. and T. Teräsvirta (1993). *Modelling Non–linear Economic Relationships.* Oxford University Press.

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." To appear, *Journal of Econometrics.*

Rissanen, J. (1987). "Stochastic complexity and the MDL principle." *Econometric Reviews* 6, 85–102.

Sin, C.–Y. and H. White (1992). "Testing for functional specification in time series: a robust approach." Working Paper, Economics, UCSD.

Teräsvirta, T. (1990). "Power properties of linearity tests for time series." Discussion Paper 90–15, Dept. of Economics, Univ. of California, San Diego.

Teräsvirta, T., D. Tjostheim and C.W.J. Granger (1992). "Aspects of modelling non–linear time series." Draft chapter for new volume of *Handbook of Econometrics*, Elsevier Science Publishers, Amsterdam.