

Chapter 1

Nonlinear Nonstationary Model Building by Genetic Algorithms

Francesco Battaglia and Mattheos K. Protopapas

Abstract Many time series exhibit both nonlinearity and nonstationarity. Though both features have been often taken into account separately, few attempts have been proposed for modeling them simultaneously. We consider threshold models, and present a general model allowing for several different regimes both in time and in levels, where regime transitions may happen according to self-exciting, or smoothly varying, or piecewise linear threshold modeling. Since fitting such a model involves the choice of a large number of structural parameters, we propose a procedure based on genetic algorithms, evaluating models by means of a generalized identification criterion. The proposed model building strategy is applied to a financial index.

Key words: time series, evolutionary computation, threshold model

1.1 Introduction

Traditional linear models such as ARMA models, cannot fit successfully nonlinear or nonstationary time series. An interesting and simple alternative might be multi-regime threshold models. A time series that follows such a model, is generated by several alternative linear autoregressive equations (the regimes). A generating process switches from one regime to another, according to the value of an indicator, that may be related to time or to another time series (called the driving variable). In the models used in this paper the driving variable is a delayed value of the time series itself, as well as time. When there are different the regimes in time, we have

Francesco Battaglia
Dipartimento di Statistica, Probabilità e Statistiche Applicate, Sapienza Università di Roma,
e-mail: Francesco.Battaglia@uniroma1.it

Mattheos K. Protopapas
Dipartimento di Statistica, Probabilità e Statistiche Applicate, Sapienza Università di Roma,
e-mail: Mattheos.Protopapas@gmail.com

a nonstationary but linear model (also called structural change, for a description see Bai and Perron, 1998; Lin and Teräsvirta, 1994). When a delayed observation determines the regime, the model is nonlinear but stationary, under suitable choices of the parameter values. These are usually called threshold models, see Tong (1990).

It has been pointed out that a complex dynamic behavior could be reproduced by a nonstationary as well as a nonlinear model, and often a confusion may arise between the two kinds of models (see Koop and Potter, 2001; Carrasco, 2002; Dupleich Ulloa, 2005). This suggests that a parsimonious model might be obtained allowing simultaneous non linearity and non stationarity. In this paper we try to fit models that have different regimes, both in time and according to a delayed observation of the time series, as the driving variable. Lundberg et al. (2003) proposed a model based on an AR(1) structure with the parameter changing both according to time and to a driving variable alternating between two regimes only, with change determined by the logistic function (in the same way as smooth transition autoregressive models of Teräsvirta, 1994). Battaglia and Protopapas (2009) introduced a change of the autoregressive parameters in a piecewise linear fashion (piecewise linear threshold multi-regime model, Baragona et al., 2004), as well as a smooth transition, and proposed a genetic algorithm to build a two regime model as well. In Battaglia and Protopapas (2010), a different structure of the chromosome encoding, made possible the introduction of a new genetic algorithm to allow for up to four regimes (in both time and levels). In this paper we give a short description of that construction, and apply the algorithm to the analysis of some financial data.

We stress our view, similar to that of Rissanen (2007), that models are only partial tools for describing some features of the data, and our aim is to discover not the absolutely correct generating process, but a reasonably parsimonious model, satisfactorily fitting the time series; we shall explain how genetic algorithms may be helpful for that search.

1.2 The Model and Estimates

The original autoregressive threshold model proposed by Tong (1990), has at each t an autoregressive structure, where parameters change according to the value of another series (the driving variable). If the driving variable is given by the delayed series itself, we have a self exciting threshold model (SETAR):

$$X_t = \phi_1^{(k)} X_{t-1} + \dots + \phi_p^{(k)} X_{t-p} + \varepsilon_t, \quad X_{t-d} \in R_k \quad (1.1)$$

where $\{R_k\}$ is a partition of the real line, and d is called the delay. Teräsvirta (1994) generalized the SETAR model to the smooth transition autoregressive model (STAR) model, in which the parameters change in a continuous, smooth way, since the transition from one regime to the next is driven by a continuous function (usually a logistic). If r_L denotes the number of regimes and $R_k = (l_{k-1}, l_k], k = 1, \dots, r_L$, the STAR equation may be written:

$$X_t = \sum_{k=1}^{r_L} \sum_{j=1}^p \phi_j^{(k)} G_{k-1}(X_{t-d}) X_{t-j} + \varepsilon_t \quad (1.2)$$

where $G_0(x) = 1$ and $G_k(x) = [1 + \exp(-\gamma_L(x - l_k))]^{-1}$. The autoregressive coefficients are essentially constant in each regime, with a continuous smooth change between regimes, whose speed is controlled by the constant $\gamma_L (> 0)$. The SETAR model may be interpreted as a special case of the STAR model, when γ tends to infinity.

Baragona et al. (2004) proposed the piecewise linear threshold model (PLTAR), where the autoregressive coefficients change linearly and continuously with the driving variable X_{t-d} , but with different slope in each regime, PLTAR is described by

$$X_t = \sum_{j=1}^p [\phi_j^{(0)} + \phi_j^{(1)} X_{t-d} + \sum_{k=2}^{r_L} \phi_j^{(k)} \max(0, X_{t-d} - l_{k-1})] X_{t-j} + \varepsilon_t \quad (1.3)$$

Here the autoregressive coefficient behaves like a linear spline across regimes. The PLTAR may be written in a similar fashion to the STAR letting $S_k(x) = \max(0, x - l_{k-1})$, $S_1(x) = x$, $S_0(x) = 1$ and

$$X_t = \sum_{k=1}^{r_L+1} \sum_{j=1}^p \phi_j^{(k)} S_{k-1}(X_{t-d}) X_{t-j} + \varepsilon_t \quad (1.4)$$

Note however that here there is an additional parameter (the linear term in X_{t-d}) for each lag. Therefore the sum over k ranges from 1 to the number of regimes plus 1.

If the model has non zero or varying means, intercept terms $\phi_0^{(k)}$ should be added. For the STAR case these new terms depend on the driving variable as before, while for a PLTAR the linear term $\phi_0^{(1)}$ as in (1.3) disappears because it would be undistinguishable from $\phi_d^{(0)}$ in the case $d \leq p$. Therefore a constant term ϕ_0 is added to the model equations (1.2) or (1.4):

$$\phi_0 = \sum_{k=1}^{r_L} \phi_0^{(k)} G_{k-1}(X_{t-d}) \quad \text{or} \quad \phi_0 = \phi_0^{(0)} + \sum_{k=2}^{r_L} \phi_0^{(k)} S_k(X_{t-d})$$

for the STAR and the PLTAR case respectively.

To model nonstationarity, we allow each of the coefficients $\phi_j^{(k)}$ to depend also on time, in a STAR or PLTAR way. Let r_T denote the number of regimes in time, and t_j denote the thresholds so that the regimes are defined by the partition $R'_k = (t_{k-1}, t_k]$, where $1 = t_0 < t_1 < \dots < t_{r_T} = N$ (where N is the series length).

We introduce STAR dependence on time using

$$\phi_j^{(k)} = \sum_{i=1}^{r_T} \beta_j(i, k) G'_{i-1}(t) \quad , \quad j = 1, \dots, p \quad (1.5)$$

with $G'_0(t) = 1$, $G'_i(t) = [1 + \exp(-\gamma_T(t - t_i))]^{-1}$, $i = 1, \dots, r_T$.

Alternatively, a time nonstationarity following a PLTAR structure may be defined as

$$\phi_j^{(k)} = \sum_{i=1}^{r_T+1} \beta_j(i, k) S'_{i-1}(t), \quad j = 1, \dots, p \quad (1.6)$$

where $S'_i(t) = \max(0, t - t_{i-1})$, $S'_1(t) = t$, $S'_0(t) = 1$.

After the combination of the different types of models that are driven by levels of the driving variable, and models driven by time, nine different kinds of models result: the type of model in time is either stationary, STAR or PLTAR, while the type of the model in levels is either linear, STAR or PLTAR. Denoting $r_L^* = r_L$ for STAR models in levels, $r_L^* = r_{L+1}$ for PLTAR, and $r_L^* = 1$ for linear models, and analogously $r_T^* = r_T$ for STAR models in time, $r_T^* = r_{T+1}$ for PLTAR in time, and $r_T^* = 1$ for stationary models, the total number of parameters is $r_L^* r_T^* p$ (or $r_L^* r_T^* (p + 1) - r_T$ if the intercept term is present). The model can be written compactly as a state dependent model (Priestley, 1988)

$$X_t = \sum_{j=0}^p \phi_j(t, X_{t-d}) X_{t-j} + \varepsilon_t \quad (1.7)$$

where the autoregressive functional coefficients $\phi_j(t, X_{t-d})$ are expressed as linear functions of the $\beta_j(i, k)$:

$$\phi_j(t, X_{t-d}) = \sum_{i=1}^{r_T^*} \sum_{k=1}^{r_L^*} \beta_j(i, k) c_{ik}(t, X_{t-d}) \quad (1.8)$$

The coefficients c_{ik} are different for different model types, and define the dependence on the transition functions; they are easily computed. Since the final model equation is linear in the elementary parameters $\beta_j(i, k)$, the β 's may be estimated by least squares.

The identification of a multi-regime model requires the selection of the solution that possesses the best properties, out of a large and discrete space of elements. Such kind of problems, common in statistical applications, have been recently addressed by means of meta-heuristic methods, and in particular genetic algorithms. In addition to several statistical identification problems, e.g. autoregressive moving average model fitting (Gaetan, 2000), outlier detection (Crawford and Wainwright, 1995; Baragona et al., 2001), variable selection in regression (Chatterjee et al., 1996), the genetic algorithm was suggested specifically for threshold models by Wu and Chang (2002); Davis et al. (2006); Battaglia and Protopapas (2009).

1.3 Genetic Algorithms

Genetic algorithms are heuristics that aim at finding near-optimal solutions in optimization problems. They are called that way because they are inspired by natural

evolution (Holland, 1975). Potential solutions of the problem are encoded as chromosomes (strings of binary digits). The algorithm works on a set of chromosomes called “population”; it evaluates these candidate solutions using a “fitness function”, and evolves them using transformation operators: crossover –where two chromosomes of the population called “parent chromosomes” are recombined to form two “children chromosomes”–, and mutation –where parts of a single chromosome are altered in a random way–, in steps called generations, to find a near optimal solution to the problem at hand. For an introduction to genetic algorithms, the reader is referred to Holland (1975) or Goldberg (1989).

1.3.1 Chromosome Encoding

In the multi-regime threshold models analyzed here, we have the following decision variables: the type of model in time and in levels, the order, and in the case of STAR and PLTAR models, the number of regimes (to a maximum of 4) in time and in levels, the delay of the driving variable and the thresholds in time and in levels. Finally, there are also the γ coefficients in time and/or in levels, in the case of a STAR model.

All the relevant decision variables hold their own part of the chromosome. These binary sub-strings are called “genes” and each of them corresponds to a value of the respective decision variable. To map the binary values of a gene to the decision variables, appropriate transformations are used; to map gene values to integer values the transformation used is:

$$x = \sum_{n=1}^L e_n 2^{n-1} + \delta \quad (1.9)$$

where x is the value of the decision variable, L is the length of the corresponding bit-field in the chromosome, e_n , the n_{th} bit of the bit-field, and δ the minimum value of the decision variable (so, as described below, for the model type $\delta = 0$, while in the case of the number of regimes $\delta = 1$). For determining the threshold values, the binary values of the corresponding genes should be transformed to real values between 0 and 1, and a different transformation is used:

$$x = \frac{\sum_{n=1}^L e_n 2^{n-1}}{2^L - 1} \quad (1.10)$$

The first two genes determine the model type. There are 3 possible types of models (STAR, PLTAR and stationary or linear), so two bits are required for each of these genes. A value of one in the first gene selects to a STAR model in time, 2 leads to a PLTAR model, and a zero or three implies a stationary model in time. The same values have the same meaning for the type of the model in levels (with the exception that we have a linear model in levels instead of a stationary model as in time); the second gene determines the type in levels.

The next genes have to do with the order of the model. The order of the model ranges from 0 to 7, and 3 bits are required. Bits 5-7 of the chromosome account for the gene that encodes the order.

The next two genes in the chromosome determine the number of regimes in time and in levels, respectively. We model a maximum of four regimes in this study, so two bits are required for each of these genes.

Coding the thresholds is a critical point, since most intuitive methods lead to many illegal chromosomes and/or to large redundancy. We use for each threshold a gene consisting of 12 bits, which is first transformed to a real value $u_i, 0 \leq u_i \leq 1$ by (1.10). Then the range of possible values for the threshold is computed and the gene u_i is mapped to that range. Since we assume a minimum of m observations in each regime, the allowable value for each threshold depends upon the number of regimes and the previous thresholds. For example, for 2 regimes in time, the first m observations should fall in the first regime, and the last m into the second; thus a value of zero in u_1 corresponds to a threshold given by the m_{th} observation of the time series and a value of one corresponds to observation $N - m$ (N is the total number of observations). If there are three regimes, at least the last m values should fall in the third regime, and at least m preceding observations should fall into the second. Then, $u_1 = 0$ determines the first threshold equal to m as before, while a value $u_1 = 1$ corresponds to the fact that the first regime is as large as possible: in that case the first threshold corresponds to observation $N - 2m$, the second regime consists of the observations $N - 2m + 1, \dots, N - m$, and the third of the observations $N - m + 1, \dots, N$: essentially, we map the interval $(0, 1)$ to the interval $(m, N - 2m)$ of admissible values for the first threshold. Finally, if there are four regimes, we must account for a minimum of m observations in each regime, therefore the admissible values for the first thresholds are between m and $N - 3m$. A similar technique is applied to the coding of the second and third threshold. The whole process may be described in the following equations.

- Two regimes. If the gene is denoted by u_1 , the threshold is $t_1 = m + (N - 2m)u_1$
- Three regimes. Genes u_1 and u_2 . Thresholds: $t_1 = m + (N - 3m)u_1$, and $t_2 = t_1 + m + (N - 2m - t_1)u_2$.
- Four regimes. The genes are denoted by u_1, u_2, u_3 . The thresholds are obtained from: $t_1 = m + (N - 4m)u_1$; $t_2 = t_1 + m + (N - 3m - t_1)u_2$, and $t_3 = t_2 + m + (N - 2m - t_2)u_3$.

The values t_i determine the values of the thresholds in time (a number between 1 and N). For thresholds in levels, an extra step is required because the thresholds are the actual observations themselves. If the decoded gene value is τ , the corresponding threshold is the observed value Y_τ , where $\{Y_t\}$ denotes the observations arranged in increasing order (this definition has essentially the same effect as if choosing any real number in-between two consecutive -in terms of magnitude- observations, as the threshold).

In our implementation we set the minimum number of observations per regime equal to $m = N/5$. The choice of m is important since it balances between the flexibility of the model and the variability in parameter estimation.

The next gene determines the delay of the driving variable in the cases of STAR or PLTAR models, while its value is not used if the chromosome represents a stationary model. It occupies 3 bits, because we assume a maximum delay of 8.

The final two genes encode the γ values for a STAR model in time and/or in delay, and consist of $L_\gamma = 7$ bits each. First a mapping of these binary numbers to real numbers between (γ_1, γ_2) , is applied:

$$\gamma = \gamma_1 + x(\gamma_2 - \gamma_1)/(2^L - 1) \quad (1.11)$$

The γ parameter controls the speed of change from 0 to 1 in the logistic function. We select a maximum value for it, that makes the STAR model essentially indistinguishable from a SETAR, i.e. such that, for a sufficiently small value ε the logistic function has value ε at the time immediately before the threshold, and value $(1 - \varepsilon)$ immediately after; and a minimum value, such that the complete variation from 0 to 1 of the logistic function requires an interval smaller than N/q . In this study we have selected $\varepsilon = 0.01$ and $q = 10$ (for details about the argument supporting this decision, see Battaglia and Protopapas, 2010).

1.3.2 Fitness, crossover and mutation

In order to compute the fitness, for each given chromosome the model is estimated by least squares and the residual variance estimate $\hat{\sigma}^2$ is computed. We use an identification criterion (generalized AIC) to determine the fitness. This is equivalent to computing a penalized gaussian likelihood: if model M has p parameters, the quantities

$$IC(M) = N \log \hat{\sigma}_M^2 + c(p)$$

are evaluated and the model with minimum value is selected. Choosing a linear penalization function $c(p) = cp$ corresponds to the generalized AIC criterion (Bhansali and Downham, 1977): the original Akaike's criterion is obtained for $c = 2$, while other criteria correspond to other choices of c ; the Schwartz criterion for example corresponds to $c = \log N$. The value $c = 3$ is selected here, as in Battaglia and Protopapas (2009), because in this case its behavior is equivalent to a test of linearity-stationarity against two-regimes alternatives proposed by Lundberg et al. (2003).

The fitness function must be related to the identification criterion through a monotone decreasing transformation, and has to be always positive. We use a simple negative exponential transformation:

$$\text{fitness}(M) = \exp\{-IC(M)/N\} = \hat{\sigma}_M^{-2} \exp\{-cp/N\}.$$

We put p equal to the total number of estimated parameters, i. e., the β 's in (1.5) and (1.6), plus the number of thresholds and gammas: thus, increasing the number

of regimes implies multiplication of the value of p , which makes our procedure relatively conservative against nonlinearity and nonstationarity.

The rule used for parent selection is “roulette wheel selection”, i.e. the probability of a chromosome to be selected as a parent is proportional to its fitness.

The crossover operator used is random point crossover (Goldberg, 1989); potential cutting points are the boundary points between genes. Consequently, genes of the children chromosomes are inherited from only one parent. After crossover, a bit-wise independent mutation operator is used to – possibly – alter binary digits of the children chromosomes; the probability of mutating a bit (from 0 to 1, or vice-versa) is selected to be 0.025. We adopted a population size equal to 50.

Finally, a form of elitism is employed: after the next generation’s population is formed, the best chromosome of the previous generation is implanted as it is into the population, thus replacing a random chromosome of the new generation’s population.

1.4 Application: Dow Jones index

We consider the daily closure values of the Dow Jones Industrial Average Index (ΔDJI) for years 2005–2009. Figure 1.1(a) displays the data, and figure 1.1(b) the returns. We only use data from January, 2005 to September, 2009 for the model (1195 observations), leaving the last three months for out-of-sample forecasts.

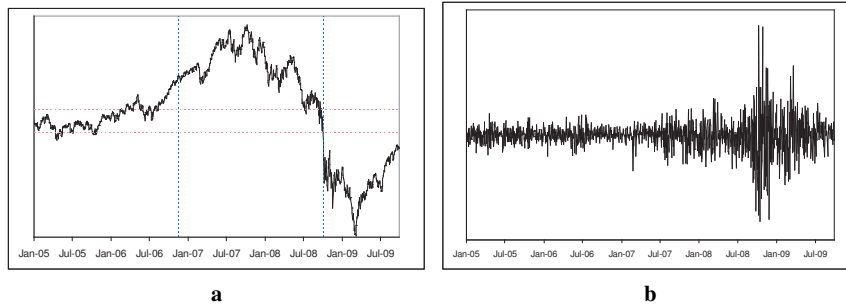


Fig. 1.1 Dow Jones Industrial Average Index, daily 2005–2009. **a.** values, **b.** returns

The series shows an apparent instability both in mean and in variance, and is not well fitted by linear models. The best ARIMA model according to the Schwarz criterion is an $ARI(2,1)$ but its residual variance, 18895, is nearly equal to that of the first difference series (19402). This corroborates the position of several analysts, that the daily variations are essentially independent, and the best forecast of tomorrow is today’s value (the *random walk* hypothesis).

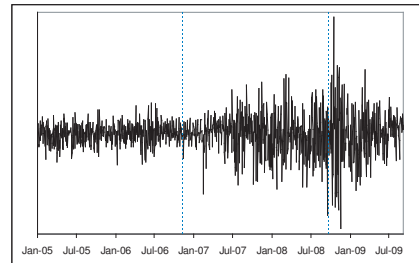
We applied our genetic algorithm- based procedure using alternatively the original values or their first differences as driving variable, and repeated the applications on the logs, transforming back to the original scale using the correction factor of Guerrero (1993). The best fitting results were obtained for original values, and will be exposed here.

The resulting optimal model is nonstationary and nonlinear, of the smooth transition type in both domains, with 3 regimes in levels (and thresholds 10252, 11090) and also 3 regimes in time (thresholds 478 and 948), order 2, delay 4 and residual variance 16111. The regimes are delimited by dotted lines in figure 1.1. We also recorded the best found linear nonstationary, and the best found nonlinear stationary, models, but their residual variance (18668 and 18428 respectively) are similar to that of the ARI(2,1) model, indicating that both features should be taken into account. However, looking at figure 1.1 and at table 1.1, where the number of observations falling into each combination of regimes in time and levels are reported, an interesting fact arises: the regimes in time and levels interact, so that essentially in the second and third time regime the data are contained in only one level regime. Thus, the model suggests that for the first time period the series follows essentially

Table 1.1 Number of observations in each regime

level	< 10252	10252-11090	> 11090
$t < 478$	19	300	151
$479 < t < 948$	2	15	453
$949 < t < 1195$	247	0	0

Fig. 1.2 Residuals of the STAR–STAR model



a nonlinear two-regime model with threshold 11090, while in the second and third period the fitting model is essentially a linear autoregressive model. The estimated model equations are (after a slight simplification) as follows:

- for the first time regime (January 2005 to November 22nd, 2006)

$$\nabla X_t = \{-0.03\nabla X_{t-1} - 70.5\}\{1 - G(X_{t-4} - 11090)\} + \{0.05\nabla X_{t-1} - 223.5\}G(X_{t-4} - 11090) + \varepsilon_t$$

- for the second time period (November 24th, 2006 to October 7th, 2008)

$$X_t = 0.86X_{t-1} + 0.125X_{t-2} + 181.5 + \varepsilon_t$$

- for the third time period (October 8th, 2008 to September 30th, 2009)

$$X_t = 0.815X_{t-1} + 0.162X_{t-2} + 240.8 + \varepsilon_t$$

The models for the second and third regime are stationary autoregressive, though one root is very near to unity. The residuals are plotted in figure 1.2.

Note that though the search space allows for very complicated models, the optimal choice is relatively simple, suggesting three different time periods, and describing the first one with a two-regime STAR model, and the next two with ordinary autoregressive structures.

On considering the forecasts obtained with this model (in table 1.2 we display the mean absolute errors of lead-1 to lead-5 forecasts for the out-of-sample observations from October 1st, 2009 to December 30th, 2009) we see that there is a substantial improvement over both the random walk and the ARI(2,1) model fitted to the entire series, and the advantage increases for larger lead times.

Table 1.2 Mean absolute errors of the lead-1 to lead-5 forecasts obtained with the ARI(2,1) model, the STAR–STAR fitted model and the random walk; observations from October, 1st to December 30h, 2009

model	lead-1	lead-2	lead-3	lead-4	lead-5
STAR–STAR	67.3	94.1	108.8	116.9	124.2
Random Walk	70.4	101.6	120.6	132.7	143.1
ARI(2,1)	75.6	108.4	128.3	147.4	165.8

It may be observed that one of the most interesting features of the financial indexes is heteroskedasticity, and also the present data, as may be seen from the returns plot in figure 1.1(b), exhibit an unstable volatility. We suggest that our procedure may also contribute to the analysis of heteroskedasticity: if the residual series of the three time regimes of the fitted model are considered separately, and an Arch-LM test performed on each of them, only for the third series the test is significant; moreover, also the Ljung-Box test does not reject the null hypothesis for the first and the second time regime. Therefore we are led to conclude that a relevant heteroskedasticity is present only in the more recent observations (beginning from October, 2008). It may be checked that this last residual sub-series is satisfactorily fitted by a GARCH(1,1), or even an ARCH(2) model.

Some other interesting features concerning volatility may be examined considering the series of the square residuals. First of all, for checking the asymmetric

volatility hypothesis, implying that the conditional variance follows two different models according to the sign of the residual, we applied the proposed procedure to the squared residuals, using the actual residuals as driving variable. The best obtained model is a STAR stationary model, with 3 regimes and thresholds -77 and -12 . It suggests that the square residuals are best fitted by different parameters for positive and negative residuals (and a further different behavior for very small negative residuals) but the variance explained by this model is only about 8%, therefore the volatility asymmetry is not so evident.

In addition, it may be of interest to check if the segmentation in time and level regimes, found for the data, is relevant also for volatility changes. To see this, we have applied our procedure to the series of the squared residuals, using the original series as driving variable. The selected model has a STAR structure in time and in levels, and explains nearly one half of the variance. It has three regimes in time and thresholds 475, 948, and three regimes in levels with thresholds 9875 and 11479. This segmentation is very similar to that obtained for the original Dow Jones series, suggesting that the identified regimes are relevant both for the index values and for its volatility.

Acknowledgements This work was supported by European Commission through Marie Curie Research and Training Network COMISEF Computational Methods in Statistics, Econometrics and Finance, and by Italian Ministry of Education through a national research grant PRIN2007.

References

- Bai, J., Perron P.: Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* **66**, 47–78 (1998)
- Baragona, R., Battaglia, F., Calzini, C.: Genetic Algorithms for the Identification of Additive and Innovation Outliers in Time Series. *Comput. Stat. Data Anal.* **37**, 1–12 (2001)
- Baragona, R., Battaglia, F., Cucina, D.: Fitting Piecewise Linear Threshold Autoregressive Models by Means of Genetic Algorithms. *Comput. Stat. Data Anal.* **47**, 277–295 (2004)
- Battaglia, F., Protopapas, M.K.: Time-varying Multi-regime Models Fitting by Genetic Algorithms. COMISEF Working Paper Series WP09. Available at <http://www.comisef.eu/files/wps009.pdf>. Cited 20 Apr 2010
- Battaglia, F., Protopapas, M.K.: Multi-regime Models for Nonlinear Nonstationary Time Series. COMISEF Working Paper Series WP26. Available at <http://www.comisef.eu/files/wps026.pdf>. Cited 20 Apr 2010
- Bhansali, R. J., Downham, D. Y.: Some Properties of the Order of an Autoregressive Model Selected by a Generalization of Akaike's EPF Criterion. *Biometrika* **64**, 547–551 (1977)
- Carrasco, M.: Misspecified Structural Change, Thresholds and Markov-Switching Models. *J. Econometrics* **109**, 239–273 (2002)

- Chatterjee, S., Laudato, M., Lynch, L. A.: Genetic Algorithms and Their Statistical Applications: an Introduction. *Comput. Stat. Data Anal.* **22**, 633–651 (1996)
- Crawford, K. D., Wainwright, R. L.: Applying Genetic Algorithms to Outlier Detection. In: Eshelman, L. J. (Ed.) *Proceedings of the Sixth International Conference on Genetic Algorithms*, pp. 546–550. Morgan Kaufmann, San Mateo (1995)
- Davis, R., Lee, T., Rodriguez-Yam, G.: Structural Break Estimation for Nonstationary Time Series Models. *J. Am. Stat. Assoc.* **101**, 223–239 (2006)
- Dupleich Ulloa, M.R.: Testing for Breaks and Threshold Effects: A Nested and Non-Nested Approach. University of Cambridge Technical Report. Available at <http://www-cfap.jbs.cam.ac.uk/events/files/ulloa.pdf>. Cited 20 Apr 2010
- Gaetan, C.: Subset ARMA Model Identification Using Genetic Algorithms. *J. Time Ser. Anal.* **21**, 559–570 (2000)
- Goldberg, D. E.: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading (1989)
- Guerrero, V. M.: Time series analysis supported by power transformations. *J. Forecasting* **12**, 37–48 (1993)
- Holland, J. H.: *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor (1975)
- Koop, G., Potter, S.: Are Apparent Finding of Nonlinearity due to Structural Instability in Economic Time Series?. *Economet. J.* **4**, 37–55 (2001)
- Lin, C., Teräsvirta, T.: Testing the Constancy of Regression Parameters Against Continuous Structural Change. *J. Econometrics* **62**, 211–228 (1994)
- Lundberg, S., Teräsvirta, T., van Dijk, D.: Time-Varying Smooth Transition Autoregressive Models. *J. Bus. Econ. Stat.* **21**, 104–121 (2003)
- Priestley, M. B.: *Non-linear and Non-stationary Time Series Analysis*. Academic Press, London (1998)
- Rissanen, J.: *Information and Complexity in Statistical Models*. Springer, Berlin (2007)
- Teräsvirta, T.: Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models. *J. Am. Stat. Assoc.* **89**, 208–218 (1994)
- Tong, H.: *Non Linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford (1990)
- Wu, B., Chang, C.-L.: Using Genetic Algorithms to Parameters (d,r) Estimation for Threshold Autoregressive Models. *Comput. Stat. Data An.* **38**, 315–330 (2002)