Detecting and modelling non-linearities in official time series

Michel Lubrano[1]

May 2002

Abstract

This paper reviews two types of non-linear threshold regression models: structural break models based on a time index and self exiting threshold regressions models. It investigates how they can be applied to detect and model breaks in the french consumption function and asymmetries in the US business cycle. The approach is Bayesian because in this type of models it is far better to average using numerical integration than to maximize a possibly discontinuous likelihood function. The final question which is addressed is the possible presence of spurious non-linearity and the tools which are available to detect them.

JEL Classification: C11, C12, D63, I29
Keywords: Bayesian inference, Hypothesis testing, Stochastic dominance, ranking economic departments

[1]GREQAM-CNRS, Centre de la Vieille Charité, 2 rue de la Charité, F-13002 Marseille, France; email: lubrano@ehess.cnrs-mrs.fr.

This paper is based on Chapter 8 of Bauwens, Lubrano and Richard (1999). It was prepared for the AMRADS Workshop in Time Series, Rhodes 27-28 May 2002. Support of the European Commission is gratefully acknowledged.

1 Introduction

I will base my talk on two quite widely debated empirical issues: the stability of the consumption function and asymmetries in the business cycle. Which type of models are adequate to investigate these two questions, is a common model feasible and how can we point out spurious non-linearities.

The paper present first a common non-linear model, which may look simple and which is to a certain extent at contrario more fashionable model recently emphasised in the literature.

In the linear regression model, the conditional expectation of the endogenous variable y_t is expressed as a linear combination of the elements of the exogenous or predetermined variables x_t'. The regression coefficient β is fixed over time. One particularly interesting way of introducing non-linearities is to suppose that the conditional expectation of the endogenous variable belongs to two or more regimes. In the simple two regime case, we have

$$\mathrm{E}(y_t|x_t) = \begin{cases} x_t'\,\beta_1, & \text{first regime,} \\ x_t'\,\beta_2, & \text{second regime.} \end{cases} \tag{1}$$

The choice between the two regimes is represented by a nonlinear transition function $F(\tilde{z}_t, \tilde{\theta})$ which takes values between zero and one. We can write

$$y_t = (1 - F(\tilde{z}_t, \tilde{\theta}))\, x_t'\beta_1 + F(\tilde{z}_t, \tilde{\theta})\, x_t'\beta_2 + \epsilon_t, \quad \mathrm{Var}(\epsilon_t) = \sigma^2, \tag{2}$$

with ϵ_t being normally distributed with zero mean and constant variance σ^2 over the two regimes. Extension to unequal variances would mean

$$\begin{aligned} y_t &= [1 - F(\tilde{z}_t, \tilde{\theta})]x_t'\beta_1 + F(\tilde{z}_t, \tilde{\theta})x_t'\beta_2 + \epsilon_t, \\ \mathrm{Var}(\epsilon_t) &= [1 - F(\tilde{z}_t, \tilde{\theta})]\sigma_1^2 + F(\tilde{z}_t, \tilde{\theta})\sigma_2^2. \end{aligned} \tag{3}$$

This model represents a transition regression model or a regression model with several regimes where the change of regimes is made according to the sign of a linear combination of the elements of $\tilde{z}_t$. In this present form, this model does not look dynamic. However, x_t and mainly z_t can and will contain dynamic variables so that this type of model is capable of producing complicated types of dynamics. It is widely used in the time series literature and the names of Tong (1983, 1990) and Teräsvirta (1994) are associated to it.

The time series literature introduced a long time ago a competing model with Lindgren (1978), model which was reconsidered more recently in the econometric literature with Hamilton (1989). This type of model is very much connected to the theory of finite mixture distributions. Basically, the endogenous variable is explained by a mixture of regression models

$$y_t = \begin{cases} x_t'\,\beta_1 + \epsilon_{1t} & \text{with probability } p, \\ x_t'\,\beta_2 + \epsilon_{2t} & \text{with probability } 1 - p \end{cases} \tag{4}$$

and can alternatively be written

$$y_t = (x_t'\beta_1 + \epsilon_{1t})s_t + (x_t'\beta_2 + \epsilon_{2t})(1 - s_t), \tag{5}$$

where s_t is a Bernouilli random variable being one with probability p and zero with probability $1 - p$. There are marked differences between (3) and (5). In model (3), the switching operates on the conditional expectation of y_t, whereas in model (5) it operates on the value of y_t itself. This makes a great difference in the likelihood function and in the identification properties. Moreover, in (3) the change of regime depends on the value of an exogenous variable z_t, whereas in (5) s_t is usually modelled as a hidden Markov chain.

Starting from (2) or (3), I have then to specify a transition function $F(.)$ and the switching variables contained in $\tilde{z}_t$. The various choices that can be made induce a large spectrum of models. First of all, $F(.)$ can be a step or a smooth function, which means that the change of regime can occur abruptly or smoothly. Second, the switching variable $\tilde{z}_t$ can simply contain a time index and a constant. The time index can be replaced by a continuous exogenous variable or a lag of the endogenous variable. Let me review some of the possible combinations which have been used in the literature:

- Combining an abrupt transition function with a time index produces the structural break model where the breaking point is unknown. The Bayesian literature on change point models is quite large and includes Ferreira (1975) and DeJong (1996) for unit roots with a structural break in the US GNP, and Lubrano (1995) for cointegrating relationships applied to a wage equation. Introducing a smooth transition function allows some delay for the change of regime to occur. This model will be used to investigate the stability of a french consumption function.

- Threshold regression models replace the time index by a continuous variable, not necessarily growing over time. This means that we can have more than one change of regime over the sample period. Bacon and Watts (1971) started a long Bayesian literature on threshold regression models which continued with e.g. Geweke and Terui (1993) and Chen and Lee (1995) and found applications in finance with e.g. Pfann, Schotman, and Tschernig (1996) or Forbes, Kalb, and Kofman (1997). This model will be used to test asymmetries in the US business cycle.

Nonlinearity has always been a delicate question for Bayesian computations, as it may imply numerical integration in high dimensions. What I show in the next sections is that the dimension of the numerical integration can be confined to one or two for most of the models I consider. This is an argument for using these models in empirical applications. Let me denote by β the vector of parameters indexing the regime equations. Conditionally on $\tilde{\theta}$, most of our models are linear in β, provided there is no element in common between β and $\tilde{\theta}$. The conditional posterior density of

β is Student provided the prior density is noninformative or naturally conjugate. The marginal posterior density of $\tilde{\theta}$, on the contrary, has to be analyzed numerically. The dimension of the numerical integration is the dimension of $\tilde{\theta}$ (minus one, because $\tilde{\theta}$ will be normalised). Posterior moments of β are found as a by-product of this integration.

The paper is organised as follows. In Section 2, I present the general Bayesian analysis of two regime models, considering explicitly the identification problem introduced by the nonlinearity and the question of the integrability of the posterior density. Unequal variances of the error term are also treated. Section 3 is devoted to testing the null hypothesis of linearity in a Bayesian framework. Section 4 analyses models where the switching depends on a time trend. In Section 5, the time trend is replaced by a continuous variable, and various types of self exciting threshold regression models are examined and their predictive densities computed. In both sections great attention is devoted to empirical applications concerning the French consumption function and the US business cycle. Section 6 details on an artifical exemple how spurious breaks can be detected. The last section concludes.

2 Bayesian analysis of the general two regime model

In this section, I propose a general Bayesian treatment for this type of model, considering first the case of equal variances, which is the most common one[1]. The case of unequal variances is treated at the end of the section. Nonlinear models display a specific identification problem when approaching linearity, which I shall solve by introducing a conditional prior information. Though the smooth transition function may appear at first sight to be more general than the step transition function, which may be viewed as a particular case, it in fact introduces a specific integrability problem which has to be solved also by the use of specific prior information. Finally, when there is a constant term in $\tilde{z}_t$, a normalisation of $\tilde{\theta}$ gives birth to a special parameter, which is called the threshold. Its posterior density may have a very peculiar shape, depending on the nature of the transition function and of the switching variable z_t.

2.1 Notation

For practical computations it is useful to reparameterise the model so as to obtain

$$y_t = x_t'\beta_1 + F(\tilde{z}_t, \tilde{\theta})\, x_t'\delta + \epsilon_t \tag{6}$$

with $\delta = \beta_2 - \beta_1$ being the contribution to the regression of considering a second regime. I shall suppose that the conditional expectation parameter β and the regime shift parameter $\tilde{\theta}$ have no element in common. Thus, it is possible to separate the parameter set into nonlinear parameters $\tilde{\theta}$ and linear parameters β. Moreover, the parameter $\tilde{\theta}$ may be partitioned into $\tilde{\theta} = [\gamma, \tilde{\theta}_2]$, where γ is a scalar parameter which

[1]For instance, Granger and Teräsvirta (1993) do not consider the threshold model with unequal variances.

measures the degree of smoothness of the transition. The change of regime itself depends on the sign of $z_t'\tilde{\theta}_2$, which shows that $\tilde{\theta}_2$ has to be normalised. In the simplest case, $\tilde{z}_t$ is composed of a variable and of a constant term, say $\tilde{z}_t = [z_t, 1]$. This implies that $\tilde{\theta}_2$ can be normalised as $\tilde{\theta}_2 = [1, -c]$. The switching is then determined by the sign of $z_t - c$, and c is referred to in the literature as the threshold parameter. Consequently the normalised nonlinear parameter is $\theta = [\gamma, c]$.

The separability between β and $\theta = [\gamma, c]$ make it useful to consider the following reparameterisation:

$$\begin{aligned} x_t'(\gamma, c) &= [x_t', F(\gamma, z_t - c)\, x_t'], \\ \beta' &= [\beta_1', \beta_2' - \beta_1'], \end{aligned} \tag{7}$$

so that the model can be written in a compact form:

$$y_t = x_t'(\gamma, c)\beta + \epsilon_t. \tag{8}$$

In the equal variance case, the variance of ϵ_t is σ^2. If I allow for different variances in each regime, I have an heteroskedastic model where

$$\mathrm{Var}(\epsilon_t) = \sigma_1^2(1 - F(.)) + \sigma_2^2 F(.). \tag{9}$$

Finally let me note that the model is not fundamentally changed if I allow for a different set of regressors x_{it} for each regime. Supposing that k_1 and k_2 are the dimensions of these sets of regressors, I shall denote by $k = k_1 + k_2$ the total number of regressors in the model. Here k_1 and k_2 are the numbers of regressors in the two regimes, though one may have $k_1 = k_2$.

2.2 Likelihood function and posterior analysis

Under a normality assumption, the likelihood function of model (8) is:

$$L(\beta, \gamma, c, \sigma^2; y) \propto \sigma^{-T} \exp\left(-\frac{1}{2\,\sigma^2}\sum_{t=1}^{T}(y_t - x_t'(\gamma, c)\beta)^2\right). \tag{10}$$

Following the distinction between linear and nonlinear parameters and the fact that in step transition models γ is dropped, it is convenient to decompose the prior into

$$\varphi(\beta, \sigma^2, \gamma, c) = \varphi(\beta, \sigma^2|\gamma)\, \varphi(\gamma)\, \varphi(c). \tag{11}$$

The class of prior densities for γ or c is not restricted, as anyway these parameters will have to be integrated out *numerically*. Moreover, I want to be free to choose the prior on c according to the particular nonlinear model which is analysed. On the contrary, it is interesting to remain in the natural conjugate framework for the prior on β (or conditionally on γ in the natural conjugate framework), as I can integrate β out *analytically*. For ease of exposition, I shall choose, for the while, a noninformative prior on β and σ^2 with

$$\varphi(\beta, \sigma^2|\gamma) \propto \sigma^{-2}. \tag{12}$$

This prior can be replaced by a normal inverted gamma2 prior without fundamentally changing the next results.

Theorem 1 *Under a noninformative prior for β and σ^2, the conditional posterior densities of β and σ^2 are given by*

$$\begin{aligned} \varphi(\beta|\gamma, c, y) &= f_t(\beta|\beta_*(\gamma, c)\,,\, M_*(\gamma, c)\,,\, s_*(\gamma, c)\,,\, T-k), \\ \varphi(\sigma^2|\gamma, c, y) &= f_{IG}(\sigma^2|s_*(\gamma, c)\,,\, T-k), \end{aligned} \tag{13}$$

where

$$\begin{aligned} M_*(\gamma, c) &= \sum x_t(\gamma, c)\, x_t'(\gamma, c), \\ \beta_*(\gamma, c) &= M_*^{-1}(\gamma, c) \sum x_t(\gamma, c)\, y_t, \\ s_*(\gamma, c) &= \sum y_t^2 - \beta_*'(\gamma, c)\, M_*(\gamma, c)\, \beta_*(\gamma, c). \end{aligned} \tag{14}$$

The corresponding posterior density of γ and c is

$$\varphi(\gamma, c|y) \propto |s_*(\gamma, c)|^{-(T-k)/2}\, |M_*(\gamma, c)|^{-1/2}\, \varphi(\gamma)\, \varphi(c). \tag{15}$$

The posterior density of γ and c has to be analysed numerically to compute its integrating constant and its moments, using a two dimensional numerical integration routine. The marginal posterior densities of β and σ^2 follow with

$$\varphi(\beta|y) = \int \varphi(\beta|\gamma, c, y)\, \varphi(\gamma, c|y)\, d\gamma\, dc, \tag{16}$$

$$\varphi(\sigma^2|y) = \int f_{IG}(\sigma^2|s_*(\gamma, c)\,,\, T-k)\, \varphi(\gamma, c|y)\, d\gamma\, dc. \tag{17}$$

The posterior density (15) has pathological behaviour in the directions of both γ and c.

2.3 Identification and nonlinearity

Model (6) becomes linear when $F(\gamma, z_t - c)$ is zero or constant. This is usually obtained for $\gamma = 0$. If $F(.)$ is the logistic distribution detailed below, the model is reduced to

$$y_t = x_t'\beta_1 + 0.5\, x_t'\delta + \epsilon_t.$$

As a consequence, the matrix $M_*(\gamma = 0, c)$ becomes singular, which means that the joint posterior density of $\theta = [\gamma, c]$ is not integrable. The marginal posterior density of β is likewise not integrable, as when $\gamma = 0$, $\delta = \beta_2 - \beta_1$ may take any value and no longer enters the likelihood function. δ is not identified when $\gamma = 0$. Finally, c is also not identified, as $F(\gamma, z_t - c)$ is constant whatever the value of c when $\gamma = 0$.

The classical literature is concerned with that identification problem only when testing for linearity and not for inference. It is the problem of optimal testing when nuisance parameters are not identified under the null, treated for instance by Hansen (1996); see also the references cited therein. Hansen (1996) applied his methodology to the threshold regression model. He proposes a transformation of the threshold parameter c (his nuisance parameter) when testing for $\delta = 0$ which makes his test procedure feasible.

In the Bayesian framework, the problem is more serious and should be tackled also for inference, as Bayesians have to integrate over the whole domain of definition of γ. It is well known that numerical problems may begin well before the point zero in some samples. The Bayesian solution to this identification and numerical problem is to introduce adequate prior information. We have two choices. One may imagine an informative prior density on γ which downweights the region near $\gamma = 0$, but that is not satisfactory. Instead we devise a prior information on δ. Consider the following conditional normal prior[2] for δ:

$$\varphi(\delta|\sigma^2, \gamma) = f_N(\delta|0, \sigma^2 \exp(\gamma)\, N_0^{-1}). \tag{18}$$

This prior becomes informative when $\gamma \to 0$ and says that δ is zero in this case. This is coherent with the linearity hypothesis, which can be described either by $\gamma = 0$ or by $\delta = 0$. When $\gamma = 0$, the prior precision is equal to N_0/σ^2. The prior precision goes to zero when γ becomes positive and the sample recovers all its freedom. As I have now an informative normal prior on a part of β, the convolution formulae given in (14) have to be changed, and more precisely the expression of $|M_*(\gamma, c)|$, which becomes

$$|M_*(\gamma, c)| = M_0(\gamma) + \sum x_t(\gamma, c)x_t'(\gamma, c) \tag{19}$$

with

$$M_0(\gamma) = \begin{pmatrix} 0 & 0 \\ 0 & N_0/\exp(\gamma) \end{pmatrix}. \tag{20}$$

Usually it suffices to take N_0 as the identity matrix or a multiple of it.

Theorem 2 *The posterior density of γ and c is infinite at the point $\gamma = 0$ in the absence of specific prior information, and finite under the partially informative normal prior (18).*

Proof: In (18), $s_*(\gamma, c)$ is always positive, as it is basically a sum of squared residuals. The crucial element is $M_*(\gamma, c)$. Expanding $M_*(\gamma = 0, c)$ given in (19) gives

$$M_*(\gamma = 0, c) = \begin{pmatrix} \sum x_t x_t' & 0 \\ 0 & N_0 \end{pmatrix}.$$

This precision matrix is singular when $N_0 = 0$, and nonsingular if both $\sum x_t x_t' > 0$ and $N_0 > 0$.

□

[2]See Schotman (1994) for a slightly different type of prior in the AR(1) model given by $(1 - \rho L)(y_t - \mu) = \epsilon_t$, where a similar identification problem occurs for the mean parameter μ when $\rho = 1$.

2.4 Integrability of the posterior density of γ

A abrupt transition is obtained as the limiting case of a smooth transition for $\gamma \to \infty$. For certain sample configurations, γ is fairly large and smooth and abrupt transitions become observationally equivalent. As a consequence,

Theorem 3 *With a flat prior, the posterior density of γ is not integrable.*

Proof: For $\gamma \to \infty$, $F(\gamma, z_t - c)$ is $O(1)$ in γ. Consequently

$$\varphi(\gamma|y) = \int \varphi(\gamma, c|y)\, \varphi(c)\, dc$$

is finite and strictly positive for any positive value of γ.

□

A similar problem arises with regression models with Student errors, as underlined in Bauwens and Lubrano (1998). The conclusion is that a sufficient prior information is needed to force the posterior density to tend to zero quickly enough at its right tail in order to be integrable. The prior should at least be $O(\gamma^{1+\upsilon})$ with $\upsilon > 0$. A convenient prior is the truncated Cauchy density[3] with

$$\varphi(\gamma) = \begin{cases} (1+\gamma^2)^{-1} & \text{if } \gamma > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

The scale of γ depends on the scale of the variable z_t. Thus it is wise to standardise the difference $z_t - c$ by the standard deviation of z_t. This makes the posterior expectation of γ interpretable.

2.5 The nature of the threshold

As the change of regime depends only on the sign of $z_t - c$, the domain of variation of c is determined by the minimum and the maximum of the sample values of z_t (if z_t is of dimension one). The interpretation of c then depends on the nature of z_t. If z_t is a time index, c represents a date and is usually denoted by τ. It is by nature a discrete parameter[4]. Its posterior density is discrete with equally spaced points, and the integration over τ should be done by direct summation. On the contrary, if z_t is a continuous variable, for instance the lagged endogenous variable as in self exciting transition autoregressive (SETAR) models, then c is a continuous parameter with a posterior density which is generally continuous. However, if the transition function is a step function, the posterior density of c presents a series of discontinuities at a

[3] A flat prior on $1/\gamma$ yields $\varphi(\gamma) \propto 1/\gamma^2$. But this prior creates a singularity at $\gamma = 0$, which is a point of interest. The Cauchy prior simply translates the singularity outside of the region $\gamma > 0$.

[4] Even if the model is reparameterised by dividing z_t by the sample size T. In that case $\tau \in [0, 1]$ becomes a rational number, but still identifies a particular observation and a date in a discrete time process.

finite number of points. Let me sort the observations of the switching variable z_t. The continuous parameter c has to be compared with these discrete observations. Between two consecutive sorted values, c can vary without modifying the sample classification and consequently the likelihood function. Thus the marginalised likelihood function $l(c; y)$ in c is a step function which looks like an histogram with $T - 1$ cells. For $T \to \infty$, the length of the steps (or the cells of the histogram) tends to zero provided the distribution of z_t is absolutely continuous wrt the Lebesgue measure over the real line. In this case, the posterior density $\varphi(c|y)$ tends to a continuous density. However, there is no reason why $l(c; y)$ or $\varphi(c|y)$ should approach a nice bell–shape. A consequence is that a pure classical approach by maximum likelihood is difficult: the likelihood function is not differentiable and may be multimodal. Any classical measure of uncertainty for c seems infeasible. The Bayesian approach, on the contrary, causes no problem, as we simply have to average on the domain of variation of c.

3 Testing for linearity

The many possibilities among the different cases of nonlinearity induce some very natural questions which can be very hard to answer.

The *first* question is of course to know if a nonlinear model is necessary to give a correct account of the behaviour of the data. Some tests presented in the classical literature do not specify a precise model for the alternative (such as the RESET test, the bispectrum test, or tests based on Volterra expansions). But most of the tests, such as the ones presented for instance in Tsay (1989) or in Lin and Teräsvirta (1994), the data are supposed to follow a particular nonlinear model under the alternative. These classical tests should be more powerful, as they are based on a precise direction, even if they may be powerful against other directions.

The *second* question of interest concerns the determination of the switching variable z_t. There is no universal answer to this question. The economic problem under inspection of course gives some natural choices, which are very helpful and which the pure statistician terribly lacks (when analysing for instance the famous data sets of Wolfe's sunspot numbers and Canadian lynx captures). In the breaking point literature, for instance, z_t is a time trend. In the switching regime literature, z_t is a continuous variable which is usually suggested by economic theory. But z_t governs the switching with some delay, so that it appears as z_{t-d}, where d is called the delay parameter. It is a statistical question to determine the value of d. In the Bayesian literature (see e.g. Geweke and Terui 1993 or Koop and Potter 1995), d is treated as a parameter for which a posterior density is derived. In this case, d is an element of θ. We could also decide that it is part of the specification of the model, as for instance the lag length in autoregressive models. Then, d should be determined with an information criterion, as for instance Tong and Lim (1980) do in a classical framework. There should be no fixed doctrine in that respect.

The linear model

$$y_t = x_t'\beta + u_t \tag{22}$$

can be tested for various misspecification directions, one of which is the presence of nonlinearity in the conditional expectation. Following the tradition of the augmented regressions of Pagan (1984), reinterpreted in a Bayesian framework by Aprahamian, Lubrano, and Marimoutou (1994), I have to look for a regression direction, essentially a set of additional variables, which express nonlinearity. The linear model (22) becomes nonlinear in particular when an $x_t'\delta\, F(\gamma, z_{t-d} - c)$ term is added. A linearisation of $F(.)$ around $\gamma = 0$ yields the additional variables looked for. Luukkonen, Saikkonen, and Teräsvirta (1988) suggest using a third order Taylor expansion, which means

$$F(\gamma, z_{t-d} - c) \simeq \alpha_0 + \alpha_1 z_{t-d} + \alpha_2 z_{t-d}^2 + \alpha_3 z_{t-d}^3. \tag{23}$$

The augmented regression for detecting a nonlinear direction becomes

$$y_t = x_t'\beta + x_t'\delta_1 z_{t-d} + x_t'\delta_2 z_{t-d}^2 + x_t'\delta_3\alpha_3 z_{t-d}^3 + \epsilon_t. \tag{24}$$

The choice between the linear model (22) and the nonlinear approximation (24) is made on the basis of a Schwarz criterion. The augmented regression is built for various values of d. The value of d which minimises the Schwarz criterion is selected. This sequence of model choices constitutes a tentative search strategy for specifying a nonlinear model in a Bayesian framework. See also Teräsvirta (1994) in a classical framework.

4 Transition models based on a time index

This type of model is concerned with the analysis of structural breaks or segmented trends at an unknown point of time. The regression coefficient β changes according to an unknown date that I call τ and treat as a discrete parameter[5]. This model was first analysed by Quandt (1958) in a maximum likelihood context and gave birth to a huge literature, mainly classical, known as the change point or structural breaks literature, which can be divided mainly into two branches. Many authors are concerned with testing parameter constancy in linear (or even nonlinear) models and thus detecting model inadequacies. Examples go from Quandt (1960) to Andrews (1993) and Lin and Teräsvirta (1994). Interest may secondly lie in estimating a two regime model with a change of regime at an unknown point of time. In the classical econometric literature, Perron and Vogelsang (1992) test for unit roots in the presence of structural change occurring at an unknown date; Gregory and Hansen (1996) and Campos, Ericsson, and Hendry (1996) test for cointegration in the same context. Economists have looked for economic slowdown after the oil crisis (Ben David and Papell 1995), breaks in inequality measures (Raj and Slottje 1994), inflation rates, and the stability of money demand equations (Baba, Hendry and Starr 1992, Raj 1995, Lütkepohl 1993).

[5] This is the c of the previous section. I have adopted this notation to stress the fact that this is a discrete parameter representing a date and also to be coherent with a part of the literature concerning this model.

4.1 Step change

The change between the two regimes is made via a step transition function $ID(t-\tau)$ defined by

$$ID(t-\tau) = \begin{cases} 0 \text{ if } t \geq \tau, \\ 1 \text{ if } t < \tau. \end{cases} \tag{25}$$

$ID(.)$ is thus an indicator function (or Heaviside function). Replacing $F(\gamma, z_t - c)$ by (25) in (6), the general model becomes

$$y_t = x_t'\beta_1 + ID(t-\tau)\, x_t'\delta + \epsilon_t \tag{26}$$

and

$$x_t'(\gamma, c) = x_t'(\tau) = [x_t', \, ID(t-\tau)\, x_t']. \tag{27}$$

Of course, the parameter γ and all the problems attached to its presence disappear. The simplest case I can think of is when the change point τ is known in advance, as then (26) is simply a model with multiplicative dummy variables. Perron (1989), for instance, used this type of model to test for unit roots in the presence of a break in the constant term and/or in the trend. With an unknown τ, we have access of course to a richer class of models. A first Bayesian treatment of this model was proposed by Ferreira (1975). Ohtani (1982) extended this model to autocorrelated errors. Lubrano (1993) and DeJong (1996) test for unit roots in the presence of an endogenous break.

The prior on τ has to incorporate the fact that τ is is a discrete parameter. A noninformative prior, for instance, would mean an equal mass on each integer $1, 2, \ldots, T$. There is an identification problem in the model whenever in regime i the number of observations (T_i) is lower than the number of variables (k_i) and the prior on β_i is noninformative. The matrix $M_*(\tau)$ is singular in this case. A prior giving a zero weight to infeasible regime classifications is one way to solve the identification problem

$$\varphi(\tau) = \begin{cases} 0 & \text{if } \tau < k_1 \text{ or } \tau > T - k_2, \\ \dfrac{1}{T - k_1 - k_2} & \text{for each other point of the sample;} \end{cases} \tag{28}$$

while still being noninformative for the remaining regime classifications. We can note that the same type of problem occurs in a classical framework when Andrews (1993) tests for structural change. He recommends choosing the prior domain of τ so that $\tau/T \in [0.15, 0.85]$.

Ferreira (1975) uses another type of noninformative prior which is based on the expression of the posterior density of τ. In my notation and using (14), the two priors advocated by Ferreira are

$$\varphi(\tau) \propto [\tau\,(T-\tau)]^{1/2} \tag{29}$$

and

$$\varphi(\tau) \propto [\tau\,(T-\tau)\,M_*(\tau)]^{1/2}. \tag{30}$$

The weights in (29) could be interpreted as coming from a beta density on the interval $[0, T]$. The second prior (30) includes a term coming from the posterior density of τ [see (15)].

The general formulation of the posterior density (15) hides the fact that the posterior density of τ is now a discrete density. As a consequence τ is integrated out by a discrete summation with $d\tau = 1$

$$\varphi(\beta|y) \propto \sum_{\tau=1}^{T} \varphi(\beta|\tau, y)\,\varphi(\tau|y) \tag{31}$$

Posterior results, as reported by Ferreira for an artificial sample, do not seem to be fundamentally affected by the shape of the prior density on τ. However, using (30) gives a posterior expectation for σ^2 which is slightly lower than in the other cases. As a matter of fact, this type of prior belongs to a wider class of prior densities which are liked by some Bayesians because they have the property of suppressing unimportant factors of the likelihood which sometimes may cause trouble. Here the prior (30) has the property of cancelling out the term $|M_*(\tau)|^{-1/2}$ from the posterior density of τ, which simplifies to

$$\varphi(\tau|y) \propto s_*(\tau)^{-(T-k)/2}[\tau\,(T-\tau)]^{1/2}. \tag{32}$$

As $|M_*(\tau)|$ does not vary very much in general (see for instance Bacon and Watts (1971), there is no major problem in using such a prior and posterior for τ. But there is also no good reason to do so, as here the term $|M_*(\tau)|$ does not cause a problem (contrary to the smooth transition case detailed below).

As the marginal posterior of τ is discrete, each point of this marginal can be interpreted as the probability that τ takes a particular value. So if no break point has a probability greater than a certain level, I can conclude with a posterior odds argument that there is no dominant break in the sample. However, the argument should be refined. There is no dominant break point if no point of the posterior of τ is higher than 0.5. But a gradual transition around a particular date may be the alternative. This is the case when the posterior of τ is concentrated around a particular value with a configuration like the one for instance presented in Figure 1 below. On the contrary, it is evidence of linearity when the probability is uniformly scattered along the domain of definition of τ.

The model with a single time break generates two regimes. It can be usefully generalised to two (or m) breaks with thus three (or $m+1$) regimes. The algebra is not substantially changed. A particular three regime model may be of special interest. A structural break means that a long term relationship is modified under the influence of a shock or a change of habit, or a change in the composition of a macroeconomic aggregate. With a single break, no way is given to the model to return to its previous state of equilibrium. In a two regime model, the step function

$$ID(\tau_1, \tau_2) = ID(t - \tau_2) - ID(t - \tau_1) \text{ with } \tau_1 < \tau_2, \tag{33}$$

where $ID(.)$ is as defined in (25), is equal to zero when t is outside the interval $[\tau_1, \tau_2]$, and one in between. This step function says that one regime corresponds to the long term situation represented by $t < \tau_1$ or $t > \tau_2$ while the supplementary regime (regime 2) corresponds to the temporary disequilibrium situation which occurs between dates τ_1 and τ_2.

4.2 Smooth transition

A smooth transition between the two regimes is obtained by replacing the step indicator function (25) by a smooth, monotonically increasing, odd function $F(.)$ with $F(-\infty) = 0$ and $F(\infty) = 1$. Many types of such functions are available (see the discussion in Bacon and Watts 1971). However, $F(.)$ is usually chosen to be a cumulative distribution: the cumulative distribution of a normal as in Goldfeld and Quandt (1973),[6] or more commonly the logistic function as advocated by Teräsvirta (1994):

$$F(\gamma, t - \tau) = \frac{1}{1 + e^{-\gamma(t-\tau)}}, \tag{34}$$

or the hyperbolic tangent as in Tsurumi (1982). The parameter γ determines the degree of smoothness in the transition and is chosen to be positive in order to define the transition function uniquely. The step transition model (26) is changed to

$$y_t = x_t'\beta_1 + F(\gamma, t - \tau)x_t'\delta + \epsilon_t \tag{35}$$

and

$$x_t'(\gamma, c) = x_t'(\gamma, \tau) = [x_t', \; F(\gamma, t - \tau)\, x_t']. \tag{36}$$

There is now one extra parameter γ, which is restricted to be positive. As underlined in Section 2, this parameter, together with the form of the nonlinearity, causes identification and integrability problems which render the use of both the noninformative Cauchy prior (21) on γ and the normal prior (18) mandatory for a safe Bayesian analysis of this model.

As the transition is smooth, it is no longer possible to classify an observation as belonging to a single regime, and consequently β becomes apparently identified, whatever the value of τ. But for large values of γ, we go back to the previous configuration, so it may be safe to keep the same prior on τ as in the step transition model.

[6] They, however, make that assumption just for estimation purposes, to implement a maximum likelihood procedure, as a step function is not differentiable and thus creates problems for a numerical optimisation routine.

With this prior information, the joint posterior density of τ and γ is well defined all over the parameter space including the point $\gamma = 0$. Its expression is

$$\varphi(\gamma, \tau|y) \propto |s_*(\gamma, \tau)|^{-(T-k)/2} \, |M_*(\gamma, \tau)|^{-1/2} \, \varphi(\tau) \, \varphi(\gamma). \tag{37}$$

This density is peculiar, as it is a mixture of a discrete and a continuous density. By a numerical integration over γ, I can compute the discrete probabilities attached to each of the $T - k$ possible values of τ. By a discrete summation over τ, I get the posterior density and moments of γ. The marginal posterior density of β is obtained by the weighted sum over τ of an integral over γ

$$\varphi(\beta|y) \propto \sum_{\tau=1}^{T} \int_{\gamma>0} \varphi(\beta|\gamma, \tau, y) \, \varphi(\gamma, \tau|y) \, d\gamma. \tag{38}$$

The interpretation of the regression coefficients β may be difficult. When the change of regime is abrupt, the value of the the function $ID(t - \tau)$ is either zero or one. So the regimes are clearly defined by a zero–one rule, and β_i is concerned only with the observations belonging to regime i. When the transition is smooth, the posterior expectations of the regression coefficients computed with (38) reflect the fact that the change of regime could be very gradual. Observation of one regime influences the regression coefficient of the other regime. So it may be rather difficult to compare the inference results on β between the two models when γ is small.

The double step function (33) can be generalised to the smooth transition case with

$$F(\gamma, \tau_1, \tau_2) = \frac{1}{1 + e^{\gamma(t-\tau_1)(t-\tau_2)}}. \tag{39}$$

This transition function was suggested as a generalisation of the exponential smooth transition function by Jansen and Teräsvirta (1996).

4.3 The French consumption function

A traditional consumption function, as reported for instance in Davidson, Hendry, Sbra, and Yeo (1978), relates the logarithm of real consumption to the logarithm of real disposable income. This formulation implies that in the long run the saving ratio is constant (provided the long run elasticity equals unity). Carruth and Henley (1990) pointed out that this ratio has declined a lot in the late eighties in England, leading to the predictive failure of many the existing consumption functions.

The French economy seems to have also experienced a fall in its saving ratio; witness Figure 3, below, which displays the ratio between CP (real consumption) and RDP (real disposable income). This is the mean propensity to consume displayed for quarterly data covering the period 1963:1-1991:4, coming from the Laroque (INSEE) database. It remained fairly stationary till 1977, despite the first oil shock of 1974. As noted by Villa (1996), profits were first to be affected by the oil shock of 1974, which was transmitted only later to wages. The share of wages in real GNP rose till 1977 and eventually till 1982. Income distribution and consumption played a countercyclical role. Adjustments operated later with the conjugate effects of wage and fiscal policies.

This change in the structure of earnings is difficult to incorporate for modelling consumption. Hendry and von Ungern-Sternberg (1981) for the UK, Brodin and Nymoen (1992) for Norway, and Villa (1996) for France introduced personal wealth as an explanation for the rate of growth of consumption. A justification can be found in the life cycle theory (see e.g. Molana 1991). But the personal wealth series has a much greater variance than the usual disposable income series. Moreover, the theoretical life cycle model supposes that consumption and wealth are cointegrated (Molana 1991). This may fail to happened for some data sets (Pierse and Snell 1995). In this section, I prefer to investigate if allowing for a temporal change in the mean propensity to consume restores the stability of the traditional consumption function, and where the break occurs. If the model is expressed in logarithms, a break on the constant term is all I need.

I shall concentrate my efforts on estimating a consumption function of the following form

$$\begin{aligned}\Delta \log CP_t &= \delta + \delta_0 \; d69.2 + \gamma \Delta \log CP_{t-4} + \beta_0 \Delta \log RDP_t \\ &\quad + (\alpha - 1)[\log CP_{t-1} - \nu \log RDP_{t-1}] + \epsilon_t,\end{aligned} \tag{40}$$

where $d69.2$ is a dummy variable which takes into account the effect of the important wage increases that followed the Matignon negotiations. In this type of consumption function, the steady state means $\log CP = \log K + \nu \log RDP$, K being the mean propensity to consume. In a dynamic equilibrium where $\Delta \log RDP = g_y$ and $\Delta \log CP = g_c = \nu g_y$, the log of the mean propensity to consume is expressed as

$$\log K = \frac{\delta}{1-\alpha} + g_y \frac{\beta_0 - (1-\gamma)\nu}{1-\alpha}. \tag{41}$$

A first estimation of this equation with a noninformative prior produces a result where the long term solution does not come in. The Schwarz criterion attached to this regression is -4.926. As I suspect a break in the constant term, I add a trend, a squared trend, and a cubic trend to build an augmented regression, coping with a possible nonlinearity in that direction. This augmented regression is preferred, as it gets a Schwarz criterion of -4.931.

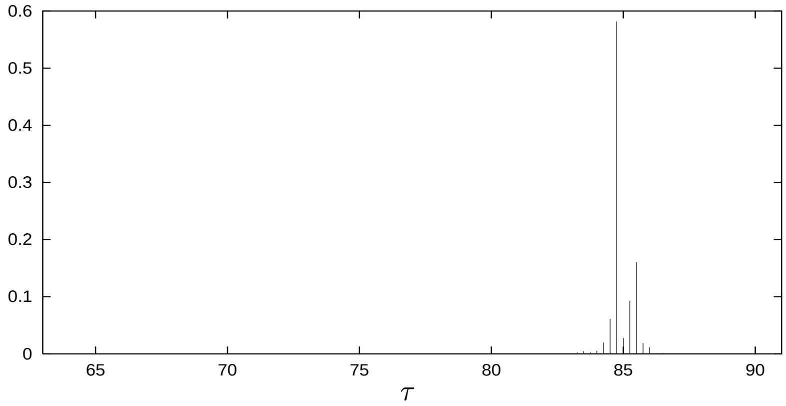
Let me now consider a step transition function for the constant term. The model is reparameterised in such a way so as to allow for the possibility of a unit long term elasticity. Using a noninformative prior, Bayesian inference produces the following output

$$\Delta \log CP_t = \underset{[0.041]}{0.11} + \underset{[0.0035]}{0.013}\, ID(t-\tau) + \underset{[0.0067]}{0.021}\, d69.2 - \underset{[0.079]}{0.23}\, \Delta \log CP_{t-4}$$
$$+ \underset{[0.067]}{0.25}\, \Delta \log RDP_t - \underset{[0.040]}{0.16} \log \left(\frac{CP}{RDP} \right)_{t-1} - \underset{[0.0031]}{0.011} \log RDP_{t-1},$$

$\mathrm{E}(\sigma^2|y) = 0.000423,$

Schwarz $= -5.030.$

Figure 1: Posterior density of τ
for the French consumption function with a step transition.

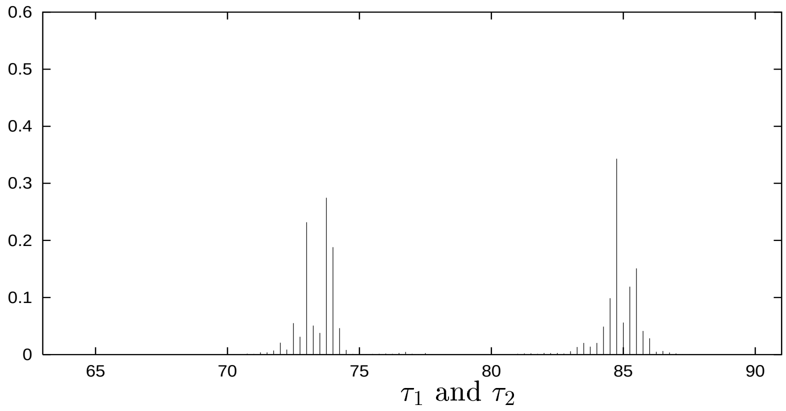


Note that this fully nonlinear model receives more evidence from the sample than the augmented regression, due to a lower value of its Schwarz criterion. The long term elasticity ν of consumption to income is lower than one, precisely 0.93 (computed with posterior means) while the marginal propensity to consume is 0.25/(1+0.23) = 0.20.

The expectation of τ is 1984.1, while the mode is 84.4. The probability of the posterior mode is 0.58. Figure 1 shows that the break point seems to be unique around this date, but 42% of the probability is scattered around this point. The shift in the constant term after 1984.1 means that the mean propensity to consume increased from $\exp[0.11/0.16 + g_x(0.25 - 0.93 \times 1.23)] = 1.99 \times \exp[-0.89 \times gx]$ to $2.16 \times \exp[-0.89 \times gx]$. These figures are unrealistic, whatever the value of g_y. They cast some doubt on the specification of the equation.

As an everlasting increase in the mean propensity to consume, as suggested by the single break model, is improbable, let me try the double break model (33), still with a diffuse prior. With this model, the unit elasticity hypothesis is accepted. The

Figure 2: Double break for the French consumption function with a step transition.



final results are

$$
\begin{aligned}
\Delta \log CP_t &= -\underset{[0.0041]}{0.0088} - \underset{[0.0017]}{0.0071} ID(\tau_1, \tau_2) + \underset{[0.0067]}{0.019}\, d692 \\
&- \underset{[0.078]}{0.26}\, \Delta \log CP_{t-4} + \underset{[0.069]}{0.23}\, \Delta \log RDP_t - \underset{[0.022]}{0.11} \log \left(\frac{CP}{RDP} \right)_{t-1},
\end{aligned}
$$

$\mathrm{E}(\sigma^2|y) = 0.000432,$

Schwarz $= -5.051.$

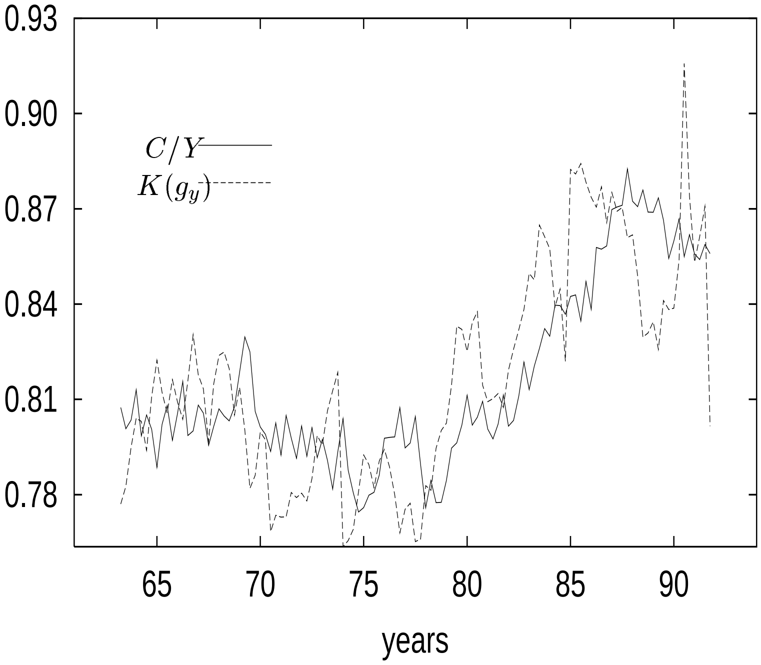
This model is slightly preferred to the one break model on the basis of the Schwarz criterion. Moreover, it has many remarkable features

- The long run elasticity is now equal to one.

- There is one break with expectation at 1984.1, as before, but also a second break earlier, in 1973.3. These breaks are well marked, with posterior standard deviations of 3 quarters.

- The mean propensity to consume is $0.92 \exp(-9.36 g_y)$ in the long term equilibrium, but seems lower, $0.87 \times \exp(-9.36 g_y)$, during the period of adjustment. With $g_y \in [0, 0.02]$, these figures are within the order of magnitude of the data. Figure 3 reproduces the data on C/Y and $K(g_y)$ with a g_y smoothed by a moving average of order 5. The two series follow similar paths. Even if the constant term is lower during the intermediate regime between 1973 and 1984, the resulting dynamic mean propensity to consume, $K(g_y)$, is slightly decreasing till 1975, but rises between 1978 and 1986, following the path shown by the ratio CP/RDP. The model thus manages to cope with the change in the structure of earning.

Figure 3: Mean propensity to consume:
data and simulation with the double threshold model.



The final question is what will yield a smooth transition function. With the conditional normal and truncated Cauchy priors, the inference results are as follows:

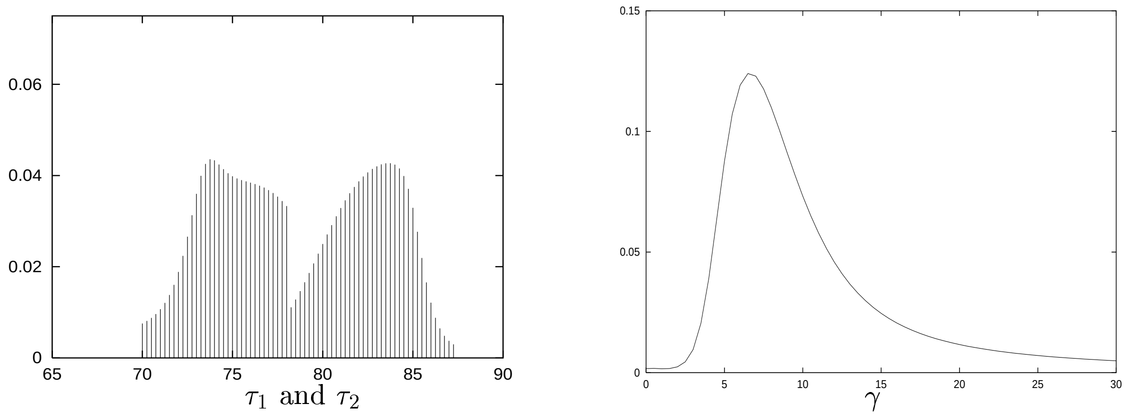
$$\begin{aligned}\Delta \log CP_t \;=\; & -\underset{[0.0047]}{0.0092} - \underset{[0.0043]}{0.012}\, F(\gamma, \tau_1, \tau_2) + \underset{[0.0069]}{0.019}\, d692 \\ & - \underset{[0.079]}{0.24}\, \Delta \log CP_{t-4} + \underset{[0.070]}{0.26}\, \Delta \log RDP_t - \underset{[0.025]}{0.10} \log \left(\frac{CP}{RDP}\right)_{t-1},\end{aligned}$$

$\mathrm{E}(\sigma^2|y) = 0.000453,$

Schwarz $= -5.003$.

The posterior densities of τ_1 and τ_2 in Figure 4 indicate the same modes for the breaking dates. But the message becomes muddled, as the probability is regularly distributed around the modes. The jump comes from the fact that the posterior densities of τ_1 and τ_2 are presented on the same graph and that they both integrate to one. The integration ranges are the same as in the step function case. The posterior density of γ indicates that high values are not rejected with a mean of 10 and a posterior 95% confidence interval of [0, 23]. As a matter of fact, the Schwarz criterion would favour the step function.

Figure 4: Double break for the French
consumption function with a smooth transition.



5 Transition models based on a continuous variable

In the previous section, the change of regime was governed by a time index, comparing t with an unknown random date τ. The switching was thus exogenous and occurred at a single point of time as z_t was trending. I now study another type of model where the switching is done by comparing either the value of the lagged endogenous variable or the value of an exogenous variable (or a linear combination of exogenous variables) with an unknown random threshold c. The switching variable(s) contained in z_t is continuous, so the threshold c is no longer a discrete parameter. But the shape of its posterior density is greatly influenced by the nature of the transition function as explained in Section 2.

The models covered in this section are very popular in statistics and econometrics. One can distinguish TR (transition regression) models with exogenous variables and TAR (transition autoregressive) models. Generally the transition is based on exogenous variables in TR models and on past values of the endogenous variable in TAR models, which thus become SETAR (self exiting transition autoregressive) models. SETAR models can be viewed as an alternative to linear ARMA models. They may introduce asymmetries in the dynamic path of the model. As the change of regime is self–exciting, these models are particularly interesting for prediction.

For these models, one has also the choice between a step and a smooth transition. Apparently some authors do prefer the step transition, such as Potter (1995) in a classical framework or Pole and Smith (1985) and Geweke and Terui (1993) in a Bayesian framework. Other authors have a marked preference for smooth transitions such as Teräsvirta (1994) (classical side) or Bacon and Watts (1971) and Peguin-Feissolle (1994) (Bayesian side). I did not manage to find in these papers a discussion concerning the choice of a type of transition, step or smooth.

The economic literature has frequently exploited the possibilities created by the use of nonlinear models. I give here some examples concerning the business cycle, the neutrality of money, and wage formation. I could have also included nonlinear taxation. As underlined first by the NBER and then by Keynes in his general theory, the dynamics of the business cycle behaves differently in expansion and in recession phases, which in general are shorter but violent. In a different domain, Keynes also stated that positive changes in the money supply are neutral, while negative changes have real effects on aggregate output. In their empirical study, Ravn and Sola (1996) found on the contrary that big changes in nominal demand are neutral, while small changes have a real effect, indicating the relevance of menu cost models and nominal rigidities. Bruno (1995) insists on nonlinear trade-off between growth and inflation. In the domain of wage formation, Johansen (1995) rejects the Phillips' curve but points out a nonlinear relation between real wages and the unemployment rate for Norway with a functional form of the type $WP = f(1/U^2)$, where WP is the level of the real wage and U the rate of unemployment. His economic model can be seen as an approximation of a truly nonlinear model. Finally, Burgess (1992) presents a model of variable adjustment costs for employment in Britain, where the adjustment cost is a function of the degree of tightness on the labour market. His model generates asymmetric dynamics.

5.1 Step transition regression models

I shall first review two step transition regression models which have been treated in the Bayesian literature and show that in fact the statistical treatment of these models is identical and enters the general framework developed in Section 2. Pole and Smith (1985) consider the following switching regression model

$$y_t = \begin{cases} x'_{1t}\beta_1 + \epsilon_t & \text{if } \tilde{z}'_t\theta \leq 0, \\ x'_{2t}\beta_2 + \epsilon_t & \text{if } \tilde{z}'_t\theta > 0. \end{cases} \tag{42}$$

Geweke and Terui (1993) and Chen and Lee (1995) consider the SETAR model of Tong and Lim (1980)

$$y_t = \begin{cases} B_1(L)\, y_t + \epsilon_t & \text{if } y_{t-d} \leq c, \\ B_2(L)\, y_t + \epsilon_t & \text{if } y_{t-d} > c. \end{cases} \tag{43}$$

These models differ by the type of switching rule. But they can be made identical if

$$\begin{aligned} \tilde{z}'_t &= [y_{t-d}, 1], \\ \theta' &= [1, -c]. \end{aligned} \tag{44}$$

Consequently a quite generic model is

$$y_t = x_t'\beta_1 + ID(z_{t-d} - c)\, x_t'(\beta_2 - \beta_1) + \epsilon_t \tag{45}$$

with

$$ID(z_{t-d} - c) = \begin{cases} 0 & \text{if } z_{t-d} - c \leq 0, \\ 1 & \text{otherwise} \end{cases} \tag{46}$$

and

$$x_t'(c) = [x_t'\,,\; ID(z_{t-d} - c)\, x_t']. \tag{47}$$

The formulation of Pole and Smith of course allows for more than one variable to govern the switching. It does not explicitly consider that there is an unknown but constant threshold c. In their model, $\tilde{z}_t$ is composed of two or more continuous and nonconstant variables (there is no constant term in their $\tilde{z}_t$). The inequality $z_t'\theta < 0$ determines cones in the θ–space when the dimension of $\tilde{z}_t$ is two, and hypercones in the general case. Some data configurations may cause problems and lead in some cases to bimodal posterior densities in small samples. For more details, see their paper. In the sequel, I shall consider only the case where $\tilde{z}_t = [z_t, 1]$, i.e. models with a constant unknown threshold c. Note however that threshold models are sometime generalised by considering more than one variable in z_t. In the simple bivariate case, the inequality $z_{1t} + \lambda z_{2t} < c$ can induce some serious pathologies for the bivariate posterior density of c and λ which are hard to understand.

The important question treated in the model of Geweke and Terui (1993) is the determination of the value of d. Geweke and Terui consider it as an extra parameter called the delay parameter and simulate the joint posterior density of (d, c) in order to get information on the optimal value of d. Note that I have advocated in Section 2 that the choice of d can also be viewed as a model selection problem which can be solved using an information criterion.

As in the case where the transition variable was a time index, the model presents an identification problem. Basically, there must be enough observations per regime. Let me order the observations in z_t by increasing values and reorder the other variables accordingly. As the threshold c is compared with the ordered values of z_t, its domain of variation is determined by the minimum and maximum values of this vector. But in order to have at least k_i observations per regime, the integration range should be limited above the k_i smallest values of z_t and below the k_i greatest values of z_t. But in case of collinearity, this is not a sufficient condition. Let us denote by H the empirical distribution of z_t. As c belongs to R, a possible prior is

$$\varphi(c) = \begin{cases} 1 & \text{if } c \in [c_a, c_b], \\ 0 & \text{otherwise} \end{cases} \tag{48}$$

with c_a and c_b such that $H(c_a) = 0.15$ and $H(c_b) = 0.85$. A similar suggestion can be found in the classical literature (Andrews 1993). Any type of informative prior could be used on the restricted domain of definition, as anyway this parameter is integrated out numerically.

Chen and Lee (1995) define a Gibbs sampler for TAR models, but the posterior analysis of this model enters the general formulation of Section 2. In the SETAR model, c is the sole nonlinear parameter, and thus inference is a one dimensional numerical integration exercise. As underlined in Section 2, the posterior density of c given by

$$\varphi(c|y) \propto |s_*(c)|^{-(T-k)/2}\,|M_*(c)|^{-1/2}\,\varphi(c) \tag{49}$$

may be extremely peculiar. It may be discontinuous for a finite number of points given by the sample values of z_t, but is continuous (and constant) between two consecutive values of z_t. Thus this parameter should be integrated out not by direct summation like τ in the previous section, but by a usual Simpson rule.

Pole and Smith (1985), Geweke and Terui (1993), and Chen and Lee (1995) consider explicitly the possibility of unequal variances between the two regimes, using the formulation given in Section 2. Various models with unequal variances are considered in Pfann, Schotman, and Tschernig (1996). It is easy to show that these models enter also the framework of Section 2. In particular, with

$$y_t = \begin{cases} x_t'\beta_1 + w_t^{\alpha}\epsilon_t & \text{if } z_t \leq c, \\ x_t'\beta_2 + w_t^{\alpha}\epsilon_t & \text{if } z_t > c, \end{cases} \tag{50}$$

the variance of the error term varies according to the variable w_t, independently of the regime switch. The authors use a Gibbs sampler to evaluate the posterior density of this model, but it can be rewritten as (45) with the variance of the error term being decomposed into

$$\mathrm{Var}(\epsilon_t) = \sigma^2 w_t^{2\,\alpha} = \sigma^2 h_t(\alpha). \tag{51}$$

The second model they consider assumes that the variance of y_t can change according to a different threshold than the conditional expectation of y_t. More precisely,

$$y_t = \begin{cases} x_t'\beta_1 + \epsilon_t & \text{if } z_t \leq c_1, \\ x_t'\beta_2 + \epsilon_t & \text{if } z_t > c_1 \end{cases} \tag{52}$$

and

$$\mathrm{Var}(\epsilon_t) = \sigma^2[(1 - ID(z_t - c_2)) + \phi\, ID(z_t - c_2)] = \sigma^2 h_t(c_2, \phi). \tag{53}$$

We have now a model with two thresholds c_1 and c_2 and a total of three nonlinear parameters. Note that c_1 and c_2 are identified by the restriction $c_1 < c_2$.

5.2 Smooth transition regression models

The step transition function is replaced in (45) by a smooth transition function. The generic model can thus be written

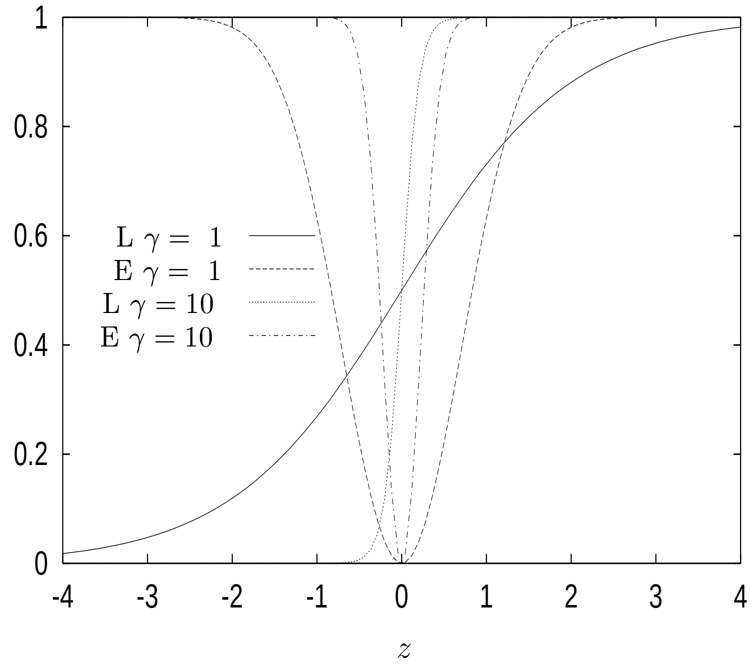
$$y_t = x_t'\beta_1 + F(\gamma, z_{t-d} - c)\, x_t'(\beta_2 - \beta_1) + \epsilon_t \tag{54}$$

and

$$x_t' = [x't\,,\ F(\gamma, z_{t-d} - c)\, x_t']. \tag{55}$$

c and γ are in general the sole nonlinear parameters of the model, leading to numerical integration of dimension two. As there is a smooth transition between the two regimes, the posterior density of c no longer presents the same discontinuities as in the step transition case.

Figure 5: Two transition functions.



The logistic transition function, already introduced in Section 4, is a generalisation of the step function (2). It is an odd function with $F(-\infty) = 0$, $F(+\infty) = 1$. In business cycle studies, it aims at detecting asymmetries between positive and negative shocks. On the contrary, an even function ($F(\pm\infty) = 1$, $F(0) = 0$), such as the exponential function used for instance by Haggan and Ozaki (1981), aims at discriminating between the effects of big shocks compared to small ones. The exponential transition function is:

$$F(\gamma, z_{t-d} - c) = 1 - e^{-\gamma\,(z_{t-d}-c)^2}. \tag{56}$$

See Bacon and Watts (1971) for other types of even functions. The difference of behaviour between the logistic and the exponential is well illustrated in Figure 5, where graphs of the logistic function and the exponential function are given for various values of γ. When $\gamma \to \infty$, the logistic function tends to the Heaviside function (46) and the model is still nonlinear. When $\gamma \to \infty$, the exponential function is always one except for the point $z_{t-d} = c$, which is of zero probability. Consequently, the model becomes linear when $\gamma = 0$, as was discussed in Sections 2 and 3, but also when $\gamma \to \infty$. The exponential transition function introduces a supplementary identification problem for $\delta = \beta_2 - \beta_1$. Note that the transition function (39), which was suggested as a generalisation of the exponential smooth transition function by Jansen and Teräsvirta (1996), does not present this problem.

Peguin-Feissolle (1994) has proposed a Bayesian analysis of the logistic smooth transition auto regressive (LSTAR) model, a simplified version of which is:

$$y_t = \beta_0 + \beta_1 y_{t-1} + F(\gamma, y_{t-d} - c)(\delta_0 + \delta_1 y_{t-1}) + \epsilon_t. \tag{57}$$

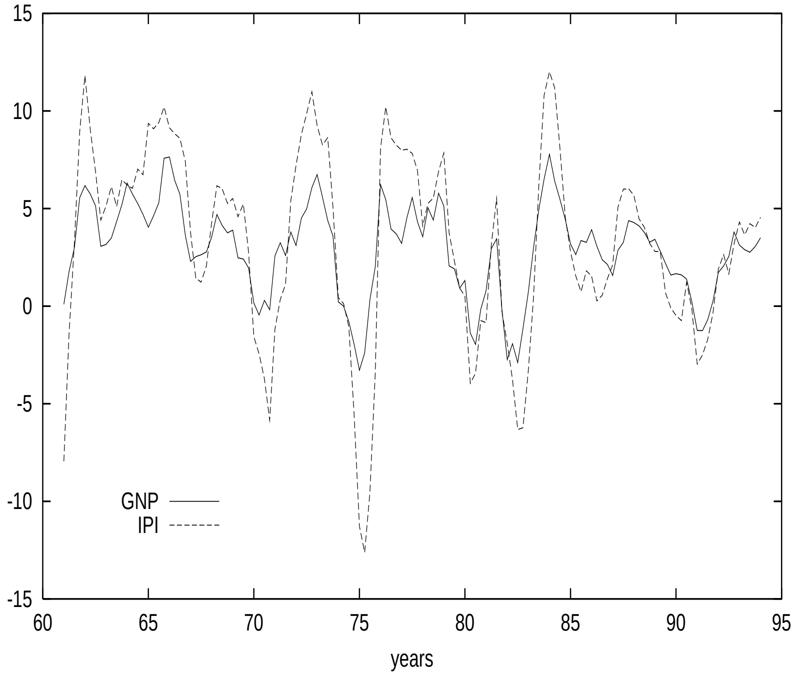
She uses Monte Carlo integration by importance sampling to analyse the posterior and predictive densities of this model. But it is easy to show that once again this model enters the general framework developed in Section 2.

> **Remark:** In the applications reported in Granger and Teräsvirta (1993, Chapter 9), it appears that the reported values of γ are fairly large for the LSTAR model, even when y_{t-d} is normalised. In the transition model based on a time index, there was presumably only one change at a particular point of time. When the transition variable is continuous and stationary, many changes of regime can occur in a given sample. All of these changes have to obey the same transition function, which has constant parameters. Consequently the threshold c is always the same and the speed of change γ is also always the same, irrespective of the size of the jump given by z_t which motivates the regime change. This restriction may explain why it is in general so difficult to make inference on γ in LSTAR models. However, when Teräsvirta and Anderson (1992) use the exponential transition function, the reported γ has midrange values, because both small and large values correspond to linearity, which is rejected by the sample.

5.3 Business cycle and nonlinear asymmetries

There is a debate about the presence or absence of nonlinearities in the US GNP. Hamilton's (1989) model explains the changes in the log of real quarterly US GNP over 1951:2 to 1984:4 using a discrete shift between two regimes, where the shift obeys a Markov process. His model reproduces the usual turning points of the US business cycle as determined by the NBER. Boldin (1996) showed that Hamilton found only a local maximum of the likelihood function and that his results were not robust to an extension of the sample period. Potter (1995), using the same data but on the period 1947:1-1990:4, adjusts a two regime step transition TAR(5) model with unequal variances where $d = 2$ and $c = 0$. Though his model seems to fit the data correctly, the linearity tests he reports, based on Hansen's (1996) theoretical results, reject linearity at 10%, but not at 5%. Hansen (1996) himself concludes a statistical artifact for nonlinearity in the US GNP. However, tests presented in Tiao and Tsay (1994) do conclude nonlinearity of the US GNP. Teräsvirta and Anderson (1992) underline that an industrial production index (IPI) has a greater variance than the GNP series and thus is more capable of displaying nonlinearity.

Figure 6: US business indicators.



I shall consider both the US real GNP and the IPI for the period 1960:1-1994:1. The data come from the OECD database. They are seasonally adjusted for the GNP. Monthly, not seasonally adjusted data are available for the IPI. I converted them to quarterly by averaging. I then took the fourth difference of these two series to stationarise them and to remove possible remaining seasonality. The resulting series are displayed in Figure 6.

I shall try first to verify the presence of nonlinearity by analysing a sequence of augmented regressions as described in Section 3. An AR(5) fits best for both series, but leads to very messy augmented regressions, as 15 supplementary regressors are needed. An AR(1) is probably biased, but nevertheless picks up a lot of the variance of the series. Let me now compare the linear AR(1) model with a sequence of augmented regressions where d varies between 1 and 5. The results are shown in Table 1. The "Schwarz" columns give the Bayesian Schwarz criterion, and the "P–value" columns the classical P–value for the F–test of linearity. At the classical 5% level, linearity is rejected for the US GNP. The smallest P–value is obtained for $d = 4$. However, a Schwarz criterion selects the linear model. If I consider an AR(5) model, linearity is also no longer rejected by classical P–values. This is in agreement with some of the empirical results mentioned above.

Table 1: Selecting linearity or nonlinearity for US GNP and US industrial production

	GNP		IPI	
d	Schwarz	P–value	Schwarz	P–value
-	-3.91	-	-2.63	-
1	-3.81	0.74	-2.52	0.89
2	-3.87	0.036	-2.62	0.004
3	-3.86	0.069	-2.66	0.001
4	-3.90	0.006	-2.55	0.25
5	-3.80	0.87	-2.53	0.68

For the US IPI, linearity is rejected at the classical level of 1%. The Schwarz criterion selects the nonlinear model with $d = 3$. This is in agreement with the results of Teräsvirta and Anderson (1992) on a smaller sample size. These results are confirmed with an AR(5).

Let me now try to adjust an LSTAR model on the IPI data only, as there is not enough evidence in favour of nonlinearity in the GNP data. Let me call the transformed variable y_t. I shall use a restricted uniform prior on c, a Cauchy prior on γ, and a conditional normal prior on the second regime regression parameters as defined in (18). I started with a model including six lags of y_t and ended up with

$$\begin{aligned} y_t \;=\;& \underset{[0.0051]}{0.016} + \underset{[0.11]}{1.06}\,y_{t-1} - \underset{[0.90]}{0.36}\,y_{t-2} - \underset{[0.19]}{1.06}\,y_{t-4} + \underset{[0.20]}{0.59}\,y_{t-5} \\ & + (1 + \exp(-\underset{[30]}{43}\; 20.34\,(y_{t-3} + \underset{[0.0020]}{0.0036})))^{-1} \\ & \times (\underset{[0.0062]}{0.020} + \underset{[0.10]}{0.25}\,y_{t-1} + \underset{[0.20]}{0.97}\,y_{t-4}) - \underset{[0.21]}{0.59}\,y_{t-5}, \\ \mathrm{E}(\sigma^2|y) =\;& 0.000245, \\ \text{Schwarz} =\;& -3.164. \end{aligned}$$

As the transition variable was normalised by its standard deviation, the posterior expectation of γ can be directly interpreted. With a value of 43, the transition between expansions and recessions is quite abrupt; witness the corresponding graph of $F(z)$ in Figure 7 which is shown together with the rescaled series z_{t-3}. Note that with a shorter sample (ending in 1986), $\mathrm{E}(\gamma|y)$ was much lower. This is in agreement with the remark made at the end of Section 5.2.

Figure 7: Transition function of the LSTAR model
for US industrial production.

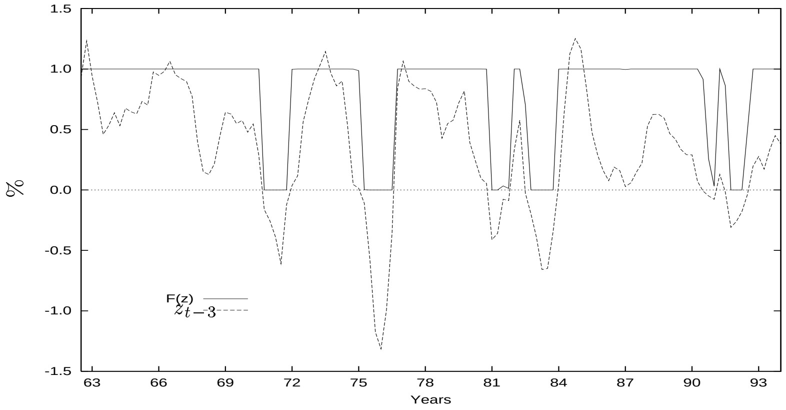
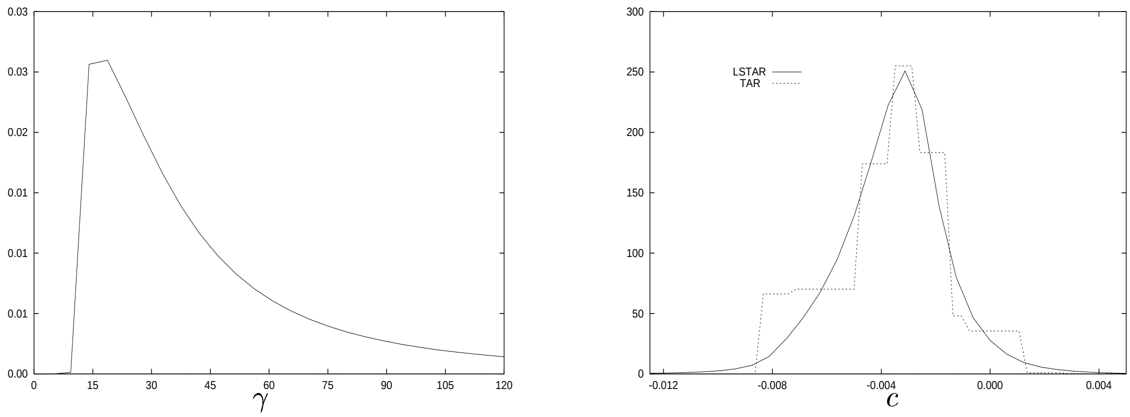


Figure 8: Posterior densities of γ and c of the LSTAR model
for US industrial production.



The posterior expectation of c is slightly negative at -0.0036, but with such a large standard deviation that it could have been taken equal to zero. The posterior density of c has not a regular shape (see Figure 8). In fact, it tends toward an irregular histogram as $\gamma \to \infty$. Note that despite the very high posterior expectation of γ, there is still a difference between the step transition and the logistic transition.

The step transition model has a Schwarz criterion of -3.165, which is only marginally better than that of the smooth transition model. There is no major change in the regression coefficients. The posterior density of the threshold is displayed in Figure 8 on the same graph as for the smooth case. It shows how the shape of this posterior density is modified by the nature of the threshold. It is less scattered, but of course presents a series of points where it is not continuous. By the way, it is still centred on zero, which suggests again that the threshold could be pinpointed at this value. In this case the model is reduced to a linear regression model with multiplicative dummy variables.

The model is characterised by very different dynamic behaviour between the two regimes. Teräsvirta and Anderson (1992) suggest computing the dominant root of the lag polynomial for each regime. I have a dominant complex root of 1.124 for the recession regime and a dominant complex root of modulus 0.827 for the expansion regime. Consequently, the dynamics of the expansion is stationary, while the dynamics of the recession is explosive, suggesting that recessions are short–lived and violent. Out of 127 observations, 25 are in the recession regime, while 102 are in the expansion regime.

6 Statistical artefacts and break detection

In this section, I would like to treat a problem which is recurrent in empirical work. What do we mean by detecting a break in an empirical equation. Has the behaviour of the agents really changed or is it just a convenient way of fitting better the data with an unachieved model? The question can be illustrated with break and unit roots. An I(1) series usually present several jumps. If the model allows for a break every time there is a jump, the I(1) property of the series disappears into the deterministic breaks and we have a statistical artefact.

6.1 A unit root model

The model I introduce now is a slight generalisation of the regression model model with a break as it allows for autocorrelated errors

$$\begin{cases} y_t &= x_t\beta_1 + ID(\tau)x_t\beta_2 + u_t \\ u_t &= \rho u_{t-1} + \epsilon_t \\ \Delta x_t &\sim N(0, v_t) \end{cases}$$

x_t is an I(1) exogenous variable and consequently y_t is also I(1). I recall that the change of regime is defined by

$$ID(\tau) = \begin{cases} 0 & \text{si } t \leq \tau \\ 1 & \text{autrement} \end{cases}$$

The error term is autocorrelated. For $\rho < 1$, the non-linear combination of of y_t and x_t is stationary, which means that these two variables are cointegrated.

6.2 Detecting artefacts

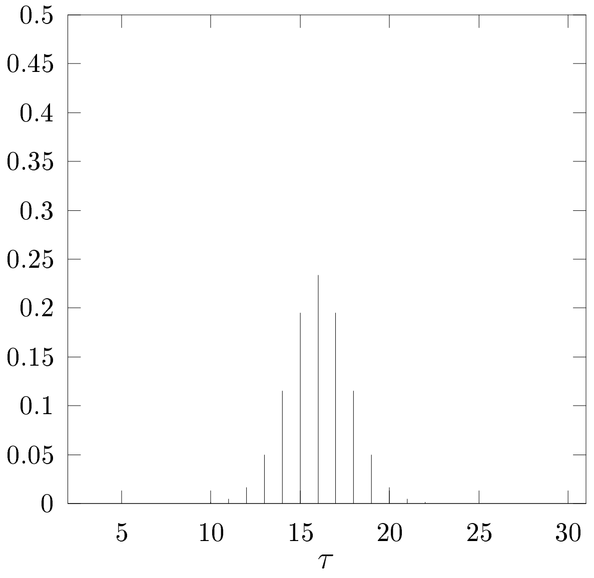
When a break is allowed for in a cointegration relationship, what is the probability of finding a spurious break in the middle of the sample as the model favours the presence of a break on a fit argument. A very simple exemple consists in regressing a trend on a constant with a break. We should find a break in the middle of the sample as in the consumption relationship. But we have diagnostic tools to show that this is in this case a real artefact.

Let us consider a trend of 32 observations and let us regress it over a constant with a break. Using a uniform prior on the coefficients, we find

$$y_t = \underset{[1.52]}{8.50} + ID(\tau)\ \underset{[1.76]}{16.00}$$

The posterior mode and expectation of τ are both equal to 16. So a break is effectively found in the middle of the sample. The first tool for detecting the misspecification

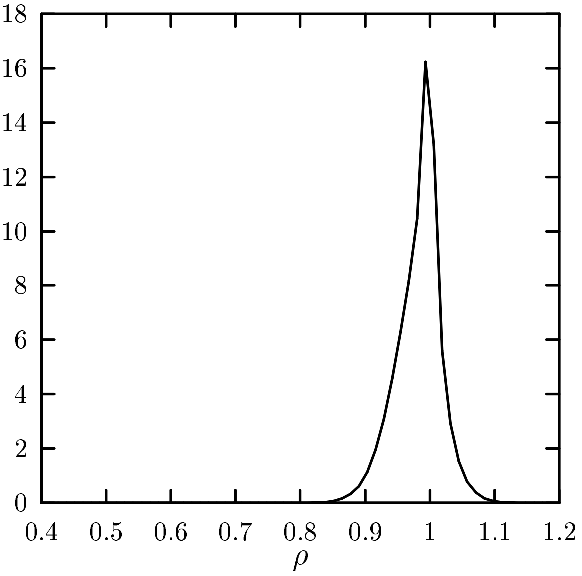
Figure 9: Posterior density of τ



consists in interpreting correctly the posterior density of ρ. It is a discrete density where each line represents the posterior probability of a break point. This it is the posterior odds of a break point compared to the whole set of all possible break points. Here the greatest mode is very small with 0.23. So there is no dominant breaking point with a posterior probability of 0.5.

The second tool is provided by the posterior density of ρ. In order to have a valid cointegration relationship, we must have $\rho < 1$. The graph of the posterior density of ρ already says us that things are going badly. If we now compute the posterior probability that $\rho > 1$, we find 0.32, whih is the proof that we have a unit root. Consequently, the model is twice badly specified and we have amply detected the statistical artefact.

Figure 10: Posterior density of ρ



7 Conclusion

Bayesian methods are flexible enough to analyse nonlinear time series models. I have shown that they lead to simple calculations requiring simple integration in one, two, eventually three dimensions. In the case of step switching (switching rules based on inequalities), Bayesian methods appear to be superior to classical ones, as they are not in trouble with the nondifferentiability of the likelihood function. Bayesian methods proceed by simple averages over the parameter space. Osiewalski and Welfe (1998), for instance, analyse a wage–price two equation model where price indexation occurs only above an unknown threshold c. Their posterior density of c has no obvious maximum. It is a collection of pieces, the shape of which is determined by the presence of a Jacobian generated by the simultaneity of the model.

The simplicity of the calculations presented in this paper may seem at odds with the modern literature on the Gibbs sampler. Whenever possible, deterministic integrations rule must be used because they are quicker and more precise than sampling methods. However, much more complicated models have been treated in the Bayesian literature, and they all need Monte Carlo methods. Lubrano (1985) analyses a switching model where β and θ are identical. This model is linked to disequilibrium models. Inference is made by importance sampling. Koop and Potter (1995) study an autoregressive threshold model where the variable governing the switching is built recursively during inference, which is conducted using the Gibbs sampler. Ghysels, McCulloch, and Tsay (1994) propose a periodic switching regime model where the change of regime follows a seasonal Markov process. This model is also estimated using the Gibbs sampler. Related papers are Albert and Chib (1993) and McCulloch and Tsay (1993).

I shall say a final word concerning smooth and step transition functions. Bayesian methods are particularly at ease with step transition functions, whereas classical methods experience most of their difficulties in this area. On the contrary, a smooth transition is amenable to a classical treatment, as its likelihood function is differentiable. But as the posterior density of this type of model is not integrable under a flat prior, a Bayesian treatment is more delicate. As a matter of fact, in the empirical examples I presented, the step transition is preferred by the data most of the time. This due to the fact that in a Bayesian framework, because of the averaging rule, a step transition function always allows for some progressivity in the transition.

Appendix A: Posterior analysis for the heteroskedastic case

.1 The heteroskedastic case

It is easy to generalise the previous model so as to allow for a different variance of the error term in each regime. I can factorise (9) into

$$\mathrm{Var}(\epsilon_t) = \sigma^2[(1 - F(\gamma, z_t - c) + \phi\, F(\gamma, z_t - c)] = \sigma^2 h_t(\gamma, c, \phi), \tag{58}$$

where $\sigma^2 = \sigma_1^2$, $\phi = \sigma_2^2/\sigma_1^2 \in [0, +\infty[$. The same analysis as before can be reproduced, provided I rescale the data by $\sqrt{h_t(\gamma, c, \phi)}$. Let me define

$$\begin{aligned} y_t(\gamma, c, \phi) &= y_t/\sqrt{h_t(\gamma, c, \phi)}, \\ x_t(\gamma, c, \phi) &= x_t(\gamma, c)/\sqrt{h_t(\gamma, c, \phi)}. \end{aligned} \tag{59}$$

Then the likelihood function of the model is

$$\begin{aligned} L(\beta, \gamma, c, \sigma^2, \phi; y) \;\propto\;& \sigma^{-T}[\textstyle\prod h_t(\gamma, c, \phi)]^{-1/2} \\ & \times\; \exp\left(-\frac{1}{2\,\sigma^2}\textstyle\sum_{t=1}^{T}(y_t(\gamma, c, \phi) - x_t'(\gamma, c, \phi)\beta)^2\right). \end{aligned} \tag{60}$$

A noninformative prior for ϕ could be

$$\varphi(\phi) \propto 1/\phi. \tag{61}$$

Theorem 4 *The joint posterior density of γ, c, and ϕ is*

$$\begin{aligned} \varphi(\gamma, c, \phi|y) \;\propto\;& [\textstyle\prod h_t(\gamma, c, \phi)]^{-1/2} \\ & \times\; |s_*(\gamma, c, \phi)|^{-(T-k)/2}\,|M_*(\gamma, c, \phi)|^{-1/2}\,\varphi(\gamma)\,\varphi(c)\,\varphi(\phi). \end{aligned} \tag{62}$$

Proof: See Appendix A.

□

Of course, compared to (15), I have increased the dimension of the numerical integration by one. Monte Carlo Markov chain algorithms may start to be competitive. The marginal posterior densities of β and σ^2 are given in Appendix A.

The natural question is now to know if an unequal variance model is really necessary. The answer is given by an inspection of the marginal posterior density of ϕ. If 1 belongs to a posterior confidence interval of ϕ, it means that that the two variances σ_1^2 and σ_2^2 are likely to be equal.

Finally, let me note that the integration problem can be simplified when the transition function is a step function, as in this case the likelihood function becomes similar to that proposed for instance in Pole and Smith (1985) or Geweke and Terui (1993). This is explained in Appendix B, where separate posterior densities for β_i and σ_i^2 are obtained. This simplification reduces the dimension of the integration problem by one, but it has a price. Separability is made possible by the fact that that conditionally on c, β_1 and β_2 are independent. But of course, they are not marginally independent. Separability implies that it is no longer possible to derive the posterior density of $\beta_2 - \beta_1$ or that of the ratio of the two error variances. This precludes some testing procedures for linearity and homoskedasticity.

Proof of Theorem 4 With (58), σ^2 can still be integrated out analytically. The conditional posterior density of β is proportional to a Student density with

$$\begin{aligned}\varphi(\beta|\gamma, c, \phi, y) \;\propto\; & [\prod h_t(\gamma, c, \phi)]^{-1/2} \\ & \times\, f_t(\beta|\beta_*(\gamma, c, \phi)\,,\, M_*(\gamma, c, \phi)\,,\, s_*(\gamma, c, \phi)\,,\, T-k),\end{aligned} \tag{63}$$

and that of σ^2 is proportional to an inverted gamma2

$$\varphi(\sigma^2|\gamma, c, \phi, y) \propto f_{IG}(\sigma^2|s_*(\gamma, c, \phi)\,,\, T-k), \tag{64}$$

where

$$\begin{aligned} M_*(\gamma, c, \phi) &= \sum x_t(\gamma, c, \phi)\, x_t'(\gamma, c, \phi), \\ \beta_*(\gamma, c, \phi) &= M_*^{-1}(\gamma, c, \phi) \sum x_t(\gamma, c, \phi)\, y_t(\gamma, c, \phi), \\ s_*(\gamma, c, \phi) &= \sum y_t(\gamma, c, \phi)^2 - \beta_*'(\gamma, c, \phi)\, M_*(\gamma, c, \phi)\, \beta_*(\gamma, c, \phi). \end{aligned} \tag{65}$$

The joint posterior density of γ, c, and ϕ is obtained as one over the integrating constant of the above Student density times its scaling factor times the prior densities of γ, c, and ϕ.

□

Corollary 1 *The marginal posterior density of* β *and* σ^2 *follows from*

$$\varphi(\beta|y) = \int f_t(\beta|\beta_*(\gamma,c,\phi)\,,\,M_*(\gamma,c,\phi)\,,\,s_*(\gamma,c,\phi)\,,\,T-k)\,\varphi(\gamma,c,\phi|y)\,d\gamma\,dc\,d\phi, \quad (66)$$

$$\varphi(\sigma^2|y) \propto \int f_{IG}(\sigma^2|s_*(\gamma,c,\phi)\,,\,T-k)\,\varphi(\gamma,c,\phi|y)\,d\gamma\,dc\,d\phi. \quad (67)$$

The latter corresponds to the posterior density of the variance of the error term in the first regime. The posterior expectation of σ^2 is obtained by taking the expectation of the analytical conditional expectation of σ^2

$$\mathrm{E}(\sigma^2|y) = \int \frac{s_*(\gamma,c,\phi)}{T-k-2}\varphi(\gamma,c,\phi|y)\,d\gamma\,dc\,d\phi. \quad (68)$$

In order to recover the posterior density of $\sigma_2^2 = \sigma^2\phi$, let me consider the one to one transformation $(\sigma^2,\phi) \longrightarrow (\sigma^2\phi,\phi)$ of Jacobian ϕ. Starting from the joint posterior density of (σ^2,ϕ) given by

$$\varphi(\sigma^2,\phi|y) \propto \int \varphi(\sigma^2|\gamma,c,\phi,y)\,\varphi(\gamma,c,\phi|y)\,d\gamma\,dc. \quad (69)$$

I get

$$\varphi(\sigma_2^2|y) \propto \int \phi\, f_{IG}(\sigma_2^2|s_*(\gamma,c,\phi)\,,\,T-k)\,\varphi(\gamma,c,\phi|y)\,d\gamma\,dc\,d\phi, \quad (70)$$

and the posterior expectation of σ_2^2 is given by

$$\mathrm{E}(\sigma_2^2|y) = \int \phi\,\frac{s_*(\gamma,c,\phi)}{T-k-2}\varphi(\gamma,c,\phi|y)\,d\gamma\,dc\,d\phi. \quad (71)$$

Appendix B: Simplification with a step transition function

The transition function $F(.)$ becomes an indicator function $ID_t(c)$ defined by

$$ID_t(c) = \begin{cases} 0 & z_t - c \geq 0, \\ 1 & \text{otherwise.} \end{cases} \quad (72)$$

Each observation is now unambiguously classified in one regime or the other. For a given sample separation (a given value for c), the conditional likelihood function is

$$\begin{aligned} L(\beta,\sigma_1^2,\sigma_2^2;y,c) \;\propto\; & \sigma_1^{-T_1(c)}\sigma_2^{-T_2(c)} \\ & \times\ \exp\left(-\frac{1}{2\,\sigma_1^2}(y_1(c)-x_1'(c)\beta_1)'(y_1(c)-x_1'(c)\beta_1)\right) \\ & \times\ \exp\left(-\frac{1}{2\,\sigma_2^2}(y_2(c)-x_2'(c)\beta_2)'(y_2(c)-x_2'(c)\beta_2)\right). \end{aligned} \quad (73)$$

The quantities $y_i(c)$, $X_i(c)$, and $T_i(c)$ are defined by

$$\begin{array}{ll} y_1(c) = [y_t(1 - ID_t(c))], & y_2(c) = [y_t ID_t(c)], \\ X_1(c) = [X_t(1 - ID_t(c))], & X_2(c) = [X_t ID_t(c)], \\ T_1(c) = \sum_{t=1}^{T}(1 - ID_t(c)), & T_2(c) = T - T_1(c). \end{array} \tag{74}$$

The subscripts indicate the regime. The symbol [.] indicates a complete matrix with generic element ".". The number of rows in $y_i(c)$ and $X_i(c)$ is equal to T, but some of the rows are rows of zeros. This type of parameterisation is different from the simple parameterisation (7) chosen in (10) and (60) and corresponds to the initial parameterisation given in (3).

The conditional posterior density of the regression parameters and of the two variances of the error terms can be decomposed into

$$\varphi(\beta, \sigma_1^2, \sigma_2^2 | c, y) \propto \prod_{i=1}^{2} f_{IG}(\sigma_i^2 | s_{*i}(c), T_i(c) - k_i)\, f_N(\beta_i | \sigma_i^2 M_{*i}^{-1}(c)) \tag{75}$$

with

$$\begin{aligned} M_{*i}(c) &= X_i'(c)\, X_i(c), \\ \beta_{*i}(c) &= M_{*i}^{-1}(c)\, X_i'(c)\, y_i(c), \\ s_{*i}(c) &= y_i'(c)\, y_i(c) - \beta_{*i}'(c)\, M_{*i}(c)\, \beta_{*i}(c). \end{aligned} \tag{76}$$

The posterior density of c is given by the integrating constants of the conditional posterior density of β, σ_1^2, and σ_2^2 times the prior density of c:

$$\varphi(c|y) \propto \varphi(c) \prod_{i=1}^{2} |s_{*i}(c)|^{-(T_i(c)-k_i)/2}\, |M_{*i}(c)|^{-1/2}. \tag{77}$$

The marginal posterior densities of β_i and σ_i^2 can all be analysed separately, and I have

$$\begin{aligned} \varphi(\beta_i|y) &\propto \int f_t(\beta_i | \beta_{*i}(c), M_{*i}(c), s_{*i}(c), T_i(c) - k_i)\, \varphi(c|y)\, dc, \\ \varphi(\sigma_i^2|y) &\propto \int f_{IG}(\sigma_i^2 | s_{*i}(c), T_i(c) - k_i)\, \varphi(c|y)\, dc. \end{aligned} \tag{78}$$

References

Albert, J., and S. Chib (1993), "Bayesian Inference of Autoregressive Time Series with Mean and Variance Subject to Markov Jumps", *Journal of Business and Economic Statistics* 11: 1–15.

Andrews, D.W.K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica* 61(4): 821–56.

Aprahamian, F., M. Lubrano, and V. Marimoutou (1994), "A Bayesian Approach to Misspecification Tests", *Document de Travail GREQAM 94A06*, Centre National de la Recheche Scientifique, Marseille.

Baba, Y., D.F. Hendry, and R.M. Starr (1992), "The Demand for Money in the U.S.A., 1960–1988, *Review of Economic Studies* 59(1): 25–61.

Bacon, D.W., and D.G. Watts (1971), "Estimating the Transition between two intersecting straight lines", *Biometrika* 58: 525–34.

Bauwens, L., and M. Lubrano (1998), "Bayesian Inference on GARCH Models using the Gibbs Sampler", *The Econometric Journal*, 1: C23–46.

Ben David, D., and D.H. Papell (1995), "The Great Wars, the Great Crash and Steady State Growth: Some New Evidence about an Old Stylized Fact", *Journal of Monetary Economics* 36(3): 453–75.

Boldin, M.D. (1996), "A Check on the Robustness of Hamilton's Markov Switching Model Approach to the Economic Analysis of the Business Cycle", *Studies in Nonlinear Dynamics and Econometrics* 1(1), 35–46.

Brodin, P.A., and R. Nymoen (1992), "Wealth Effects and Exogeneity: The Norvegian Consumption Function 1966(1)–1989(4)", Oxford Bulletin of Economics and Statistics 54(3): 431–54.

Bruno, M. (1995), *Inflation, Growth and Monetary Control: Nonlinear Lessons from Crisis and Recovery*, Bank of Italy Paolo Baffi Lectures on Money and Finance. Rome: Edizioni dell'Elefante.

Burgess, S.M. (1992), "Nonlinear Dynamics in a Structural Model of Employment", *Journal of Applied Econometrics* 7(S): 101–18.

Campos, J., N.R. Ericsson, and D.F. Hendry (1996), "Cointegration Tests in the Presence of Structural Breaks", *Journal of Econometrics* 70(1): 187–220.

Carruth, A., and A. Henley (1990), "Can Existing Consumption Functions Forecast Consumer Spending in the Late 1980's?", *Oxford Bulletin of Economics and Statistics* 52: 211–22.

Chen, C.W.S., and J.C. Lee (1995), "Bayesian Inference of Threshold Autoregressive Models", *Journal of Time Series Analysis* 16: 483–92.

Cook, P., and L.D. Broemeling (1996), "Analyzing Threshold Autoregressions with a Bayesian Approach", in T.B. Fomby (ed.) *Advances in Econometrics: Bayesian Methods Applied to Time Series Data*, Vol. 11, Part B. Greenwich: JAI Press Inc., pp. 89–107.

Davidson, J.E.H., D.F. Hendry, F. Sbra, and S. Yeo (1978), "Econometric Modelling of the Aggregate Time Series Relation between Consumption and Income in the UK", *The Economic Journal* 88: 661–92.

DeJong, D.N. (1996), "A Bayesian search for Structural Breaks in U.S. GNP", in T.B. Fomby (ed.), *Advances in Econometrics: Bayesian Methods Applied to Time Series Data*, Vol. 11, Part B. Greenwich: JAI Press Inc., pp. 109–46.

Ferreira, P.E. (1975), "A Bayesian Analysis of a switching Regression Model: Known Number of Regimes", *Journal of the American Statistical Association* 70: 370–74.

Forbes, C.S., G.R.J. Kalb, and P. Kofman (1997), "Bayesian Arbitrage Threshold Analysis", *Monash University Discussion Paper.* Clayton, Australia.

Geweke, J. (1989), "Exact Predictive Densities for Linear Models with ARCH Disturbances", *Journal of Econometrics* 40: 63–86.

Geweke, J., and N. Terui (1993), "Bayesian Threshold Autoregressive Models for Nonlinear Time Series", *Journal of Time Series Analysis* 14: 441–54.

Ghysels, E., R.E. McCulloch, and R.S. Tsay (1994), "Bayesian Inference for Periodic Regime-Switching Models", Working Paper, CRDE, Université de Montréal.

Goldfeld, S.M., and R.E. Quandt (1973), "A Markov Model for Switching Regressions", *Journal of Econometrics* 1: 3–16.

Granger, C.W.J., and T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships.* Oxford: Oxford University Press.

Gregory, A.W., and B.E. Hansen (1996), "Tests for Cointegration in Models with Regime and Trend Shifts", *Oxford Bulletin of Economics and Statistics* 58(3): 555–60.

Haggan, V., and T. OZAKI (1981), "Modelling Nonlinear Random Vibrations Using an Amplitude-Dependant Autoregressive Time Series Model", *Biometrika* 68: 189–96.

Hamilton, G.D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica* 57: 357–84.

Hansen, B.E. (1996), "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis", *Econometrica* 64 (2): 413–30.

Hendry, D.F., and T. von Ungern-Sternberg (1981), "Liquidity and Inflation Effects on Consumers' Expenditure", in A. Deaton (ed.), *Essays in the Theory and Measurement of Consumer Behaviour.* Cambridge: Cambridge University Press.

Jansen, E.S., and T. Teräsvirta (1996), "Testing Parameter Constancy and Super Exogeneity in Econometric Equations", *Oxford Bulletin of Economics and Statistics* 58(4): 735–63.

Johansen, K. (1995), "Norwegian Wage Curves", *Oxford Bulletin of Economics and Statistics* 57(2): 229–47.

Koop, G., and S.M. Potter (1995), "Bayesian Analysis of Endogenous Delay Threshold Models Using the Gibbs Sampler", Working Paper, Department of Economics, University of California at Los Angeles.

Lin, C.F., and T. Teräsvirta (1994), "Testing the Constancy of Regression Parameters against Continuous Change", *Journal of Econometrics* 62 (2): 211–28.

Lindgren, G. (1978), "Markov Regime Models for Mixed Distributions and Switching Regressions", *Scandinavian Journal of Statistics* 5: 81–91.

Lubrano, M. (1985), "Bayesian Analysis of Switching Regression Models", *Journal of Econometrics* 29: 69–95.

Lubrano, M. (1993), "Bayesian Unit Root Tests with Endogenous Break Point: An Application to the Analysis of Persistence in European Unemployment Rate Series", Mimeo, SPES Workshop on Wage and Price Formation, Maastricht, February.

Lubrano, M. (1995), "Bayesian Tests for Co-integration in the Case of Structural Breaks: An Application to the Analysis of Wage Moderation in France", *Recherches Economiques de Louvain* 61(4): 479–507.

Lütkepohl, H. (1993), "The Sources of the U.S. Money Demand Instability", *Empirical Economics*, 18(4): 729–43.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988), "Testing Linearity against Smooth Transition Autoregression", *Biometrika* 75: 491–99.

McCulloch, R.E., and R.S. Tsay (1993), "Bayesian Inference and Prediction for Mean and Variance Shifts", *Journal of the American Statistical Association* 423: 968–78.

Molana, H. (1991), "The Time Series Consumption Function: Error Correction, Random Walk and the Steady-State", *The Economic Journal* 101: 382–403.

Ohtani, K. (1982), "Bayesian Estimation of the Switching Regression Model with Autocorrelated Errors", *Journal of Econometrics* 18: 251–61.

Osiewalski, J., and A. Welfe (1998), "The Wage-Price Mechanism: An Endogenous Switching Model", *European Economic Review* 42: 365–74.

Pagan, A. (1984), "Model Evaluation by Variable Addition", in D.F. Hendry and F.K. Wallis (eds), *Econometrics and Quantitative Economics.* Oxford: Blackwell.

Peguin-Feissolle, A. (1994), "Bayesian Estimation and Forecasting in Nonlinear Models Application to an LSTAR Model", *Economic Letters* 46: 187–94.

Perron, P. (1989), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Econometrica* 57: 1361–401.

Perron, P., and T.J. Vogelsang (1992), "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity", *Journal of Business and Economic Statistics* 10(3): 301–20.

Pfann, G.A., P.C. Schotman, and R. Tschernig (1996), "Nonlinear Interest Rate Dynamics and the Implications for the Term Structure", *Journal of Econometrics* 74: 149–76.

Pierse, R.G., and A.J. Snell (1995), "Temporal Aggregation and the Power of Tests for a Unit Root", *Journal of Econometrics* 65(2): 333–45.

Pole, A.M., and A.F.M. SMITH (1985), "A Bayesian Analysis of Some Threshold Switching Models", *Journal of Econometrics* 29: 97–119.

Potter, S.M. (1995), "A Nonlinear Approach to US GNP", *Journal of Applied Econometrics* 10: 109–25.

Quandt, R.E. (1958), "The Estimation of Parameters of a Linear Regression System obeying two Separate Regimes", *Journal of the American Statistical Association* 53: 873–80.

Quandt, R.E. (1960), "Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes", *Journal of the American Statistical Association* 55: 324–30.

Raiffa, H. and R. Schlaifer (1961), "*Applied Statistical Decision Theory.* Boston: Harvard University Press.

Raj, B. (1995), "Institutional Hypothesis of the Long-Run Income Velocity of Money and Parameter Stability of the Equilibrium Relationship", *Journal of Applied Econometrics* 10(3): 233–53.

Raj, B., and D.J. Slottje (1994), "The Behaviour of Alternative Income Inequality Measures in the United States from 1947–1990 and the Structural Break", *Journal of Business and Economic Statistics* 12(4): 479–87.

Ravn, M., and M. Sola (1996), "A Reconsideration of the Empirical Evidence on the Asymmetric Effects of Money Supply Shocks: Positive versus Negative or Big versus Small?" Memo 1996/4, Dept of Economics, University of Aarhus.

Schotman, P.C. (1994), "Priors for the AR(1) Model", *Econometric Theory* 10: 579–95.

Teräsvirta, T. (1994), "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models", *Journal of the American Statistical Association* 89: 208–18.

Teräsvirta, T., and H. Anderson (1992), "Modelling Nonlinearities in Business Cycles using Smooth Transition Autoregressive Models", *Journal of Applied Econometrics* 7: S119–36.

Tiao, G.C., and R.S. Tsay (1994), "Some Advances in Nonlinear and Adaptive Modelling in Time Series", *Journal of Forecasting* 13: 109–31.

Tong, H. (1983), *Threshold Models in Nonlinear Time Series Analysis.* New-York: Springer Verlag.

Tong, H. (1990), *Nonlinear Time Series: A Dynamical System Approach.* Oxford University Press.

Tong, H., and K.S. Lim (1980), "Threshold Autoregression, Limit Cycles and Cyclical Data", *Journal of the Royal Statistical Association* B42: 245–92.

Tsay, R.S. (1989), "Testing and Modelling Threshold Autoregressive Processes", *Journal of the American Statistical Association* 84: 231–40.

Tsurumi, H, (1982), "A Bayesian and Maximum Likelihood Analysis of a Gradual Switching Regression in a Simultaneous Equation Framework", *Journal of Econometrics* 19: 165–82.

Villa, P. (1996), "La Fonction de Consommation sur Longue Période en France", it Revue Economique 47(1): 111–42.