SMOOTH TRANSITION GARCH MODELS:
A BAYESIAN PERSPECTIVE

Michel Lubrano[1]

December 1998
Revised February 2001

1 Introduction

Financial series are characterised by periods of large volatility followed by periods of relative quietness. This type of clustering led to the idea that volatility is predictable, which is of primary importance for option pricing, for portfolio selection and for designing optimal dynamic hedging strategies. The ARCH and GARCH models introduced respectively by Engle (1982) and

[1]GREQAM-CNRS, 2 rue de la Charité, 13002 Marseille, France and CORE, 34 voie du Roman Pays, B-1348 Louvain la Neuve, Belgique. **email**: lubrano@ehess.cnrs-mrs.fr

The basic idea of this paper (smooth transition GARCH) grew out during the visit of Timo Terasvirta at GREQAM in September 1996. Once a first version of this paper was completed and presented at EC^2, Florence December 1996, it appeared that a similar effort has been pursued in a classical framework by Gonzales-Riviera (1996) and by by Hagerud (1997). Some of the ideas concerning the reparameterisation of GARCH models have been explored in discussions with Rob Engle during his visit in Marseille in June 1992, conversations later pursued with Jean-François Richard. I am grateful to participants of seminars in Marseille, Bordeaux, Louvain, Maastricht, to participants to the conference EC2 "Simulation Methods in Econometrics" held in Florence, December 1996, to the Krakow Workshop on Bayesian Econometrics and Statistics, May 1997. Many discussions and critics with Luc Bauwens helped a lot in the writing of the paper. Usual disclaimers apply.

Support of the European Commission Human Capital and Mobility Program through the network "Econometric inference using simulation methods" is gratefully acknowledged. This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the author.

by Bollerslev (1986) were quite successful in predicting volatility compared to more traditional methods as underlined for instance in Engle, Kane and Noh (1996) or Noh, Engle and Kane (1995). But better predictions are obtained when asymmetries [Engle and Ng (1993)] and non-linearities [Pagan and Schwert (1990)] in the response of volatility to news arriving on the market are taken into account. The "leverage effect"[1] that commonly represents the asymmetric impact of good news and bad news on volatility has certainly be the most widely explored with Nelson (1991) who proposed his EGARCH model or Glosten, Jagannathan and Runkle (1993) (GJR from now on) and also Zakoian (1994) who proposed a threshold GARCH or TGARCH. Engle and Ng (1993) provided a summary of asymmetric GARCH models and introduced some new formulations. They tested these models to the daily return series of the Japanese TOPIX index covering 1980-1988. If on average these models perform better than the symmetric GARCH, they found evidence of mis-specification in all the models, even if the TGARCH was one of the most successful. This suggests that there is still some room for finding a flexible non-linear modelling of the skedastic function.

This paper considers a new class of GARCH models that introduces a smooth transition between two regimes defined by a threshold. I named this model STGARCH for Smooth Transition GARCH. As financial data have very often a high frequency of observation, a smooth transition seems a priori better than an abrupt transition. Engle and Ng (1993) found that the most severe mis-specification direction was that the tested models did not take adequately account of the sign asymmetry. The smooth transition model that is proposed in this paper addresses the problem of sign asymmetry. But it is more than a simple generalisation of the TGARCH as it allows for various transition functions that confer a great flexibility to the skedastic function, taking into account sign but also size effects. Finally the specification retained accepts the simple GARCH as a restriction.

The approach of the paper is Bayesian in its spirit. When a non-linearity is introduced in a model, the likelihood function becomes tricky to maximise as it may be non-differentiable. In such a case, averaging is more secure than maximising, even if the computational burden may be severe. In this paper I shall use a version of the Gibbs sampler that is the Griddy Gibbs sampler

[1]The term "leverage effect" comes from the fact that a decrease in the stock price would increase the financial leverage of the firm. This implies a negative correlation between volatility and past returns. In GARCH models, the conditional variance h_t is a function only of the squares of past errors u_{t-j}^2 and not of their signs, which precludes any correlation between h_t and u_{t-j}.

developed in Bauwens and Lubrano (1998). Specification tools appeared rapidly as a necessity because of the computation burden involved in trying to fit different types of non-linear models. A Bayesian specification search is thus developed in the paper. The strategy is based on the comparison between the posterior and the predictive variance of the Bayesian residuals. It draws on Bauwens and Lubrano (1991). The Bayesian approach provide a nice and convenient framework and the test regressions proposed are compared in the paper to some of the existing tests available in the literature.

The paper is organised as follows. In section 2, I present the general smooth transition GARCH with appropriate transition functions for sign and size asymmetries. In section 3, I compare the impact curves of these different models and stress the role of a threshold parameter . In section 4, I introduce the methodology for discriminating between different types of asymmetry. In section 5, I study Bayesian inference in GARCH and STGARCH models, insisting on various available parameterisation for the GARCH and on the fact that the posterior density of the transition parameter in STGARCH is not integrable under a flat prior. The next two sections are devoted to empirical applications concerning the Brussels and the Tokyo stock indexes. A last section concludes.

2 A general class of smooth transition GARCH models

The simple GARCH(1,1) is certainly the most widely used model for predicting the volatility of financial series [see Bollerslev, Chou and Kroner (1992) for a good review on the topic]. The regression model with GARCH(1,1) errors can be written as follows:

$$\begin{cases} y_t = x_t'\delta + u_t \\ u_t = \epsilon_t\sqrt{h_t} \qquad \epsilon_t \sim N(0,1) \\ h_t = \omega + \alpha u_{t-1}^2 + \beta h_{t-1} \end{cases} \tag{1}$$

Most of the time, x_t contains a constant and lagged values of y_t. In (1), the conditional expectation of y_t is $x_t'\delta$ and the unpredicted part of y_t is $u_t = y_t - x_t'\delta$. This represents the "news" arriving on the market as defined by Engle and Ng (1993). In the GARCH model, news have a symmetric impact on volatility, whatever their sign or magnitude and whatever the level of y_t. I shall use this model as a starting point to introduce asymmetry and level effects on volatility.

2.1 Some existing models with asymmetry

The possibility of an asymmetric impact of news on volatility has for long been suspected in the literature and confirmed among others by Nelson (1991). The EGARCH model constitutes the first introduction of an asymmetric effect between negative and positive shocks in an econometric model of volatility with:

$$\log(h_t) = \omega + \alpha g(\epsilon_{t-1}) + \beta \log(h_{t-1}) \tag{2}$$

The formulation in logarithm relaxes the usual positivity constraint on the parameters. The asymmetric effect is introduced by the non-linear function $g(\epsilon_t)$:

$$g(\epsilon_t) = \theta\, \epsilon_t + \gamma[|\epsilon_t| - \mathrm{E}(|\epsilon_t|)] \tag{3}$$

which is a function of both the magnitude and the sign of ϵ_t. By construction it has a zero mean. For positive ϵ's, $g(\epsilon_t)$ is a linear function with slope $\theta + \gamma$. For negative ϵ's, the slope becomes $\theta - \gamma$. θ allows for possible negative correlation between ϵ_t and future values of the skedastic function h_t (leverage effect). Under the normality assumption, $\mathrm{E}(|\epsilon_t|) = \sqrt{2/\pi}$.

The success of the EGARCH model motivated a large literature of follower models. See Engle and Ng (1993) for a review. The model of GJR (1993) is directly related to the present paper[2]:

$$h_t = \omega + \alpha_1 u_{t-1}^2 (1 - S_{t-1}) + \alpha_2 u_{t-1}^2 S_{t-1} + \beta h_{t-1} \tag{4}$$

S_t is an indicator function that is zero when u_t is negative and one otherwise. This formulation introduces an asymmetry of reaction for the conditional variance. The change of regime occurs when u_t crosses the threshold zero. I call this model a threshold GARCH or TGARCH[3]. Compared to the EGARCH model, moments are simpler to compute as underlined in Pagan (1996).

[2]Zakoian (1994) suggested a model equivalent to:

$$\sqrt{(h_t)} = \omega + \alpha_1 u_{t-1}(1 - S_{t-1}) + \alpha_2 u_{t-1} S_{t-1} + \beta h_{t-1}$$

The news impact curve of this model is different from that of GJR (1993) due to the squaring of h_t. It is not minimum at $u = 0$.

[3]Zakoian (1994) used TARCH for his model.

2.2 Smooth transition between negative and positive shocks

Threshold GARCH models can be generalised using a smooth transition function $F(u_{t-1},\gamma)$ taking continuous values between zero and one. The parameter γ governs the smoothness of the transition. Using this smooth transition function, the two regime skedastic function in (4) becomes:

$$\begin{aligned} h_t &= \omega + \alpha_1 \, u_{t-1}^2[1 - F(u_{t-1},\gamma)] + \alpha_2 \, u_{t-1}^2 F(u_{t-1},\gamma) + \beta \, h_{t-1} \\ &= \omega + \alpha_1 \, u_{t-1}^2 + \lambda \, u_{t-1}^2 \, F(u_{t-1},\gamma) + \beta \, h_{t-1} \end{aligned} \tag{5}$$

where $\lambda = \alpha_2 - \alpha_1$. Among the many possible odd smooth transition functions[4], the logistic function was proved to be very convenient in a classical non-linear modelling framework by Terasvirta (1994). For a smooth transition GARCH where the objective is to allow for a possible difference of reaction between negative and positive shocks, this function is:

$$F(u_{t-1},\gamma) = \frac{1}{1 + \exp(-\gamma \, u_{t-1})} \tag{6}$$

The function $F(.)$ tends to zero when $u \to -\infty$ and to one for $u \to +\infty$. So α_1 will characterise negative shocks and α_2 positive ones. For $\gamma \to \infty$, the transition function becomes equivalent to the Dirac function S_t defined above. This formulation is more flexible than GJR (1993) or Zakoian (1994) as it allows a gradual transition that may be an important feature for high frequency data. For a similar suggestion see Gonzales- Riviera (1996) or Hagerud (1997). I shall call this model LSTGARCH for logistic STGARCH.

2.3 Smooth transition between small and big shocks

Periods of important volatility do not last for long in financial series. For instance the great crash of October 1987 gave birth to a peak of the variance in the S&P500, but volatility dampened very quickly. Engle and Mustapha (1992) found a a variable persistence of shocks for the S&P500, small shocks being more persistent than big shocks. Susmel and Engle (1994) detected an symmetry between small and big shocks for the New-York and London equity market. An even[5] transition function like the exponential function

$$F(u_{t-1},\gamma) = 1 - \exp(-\gamma \, u_{t-1}^2) \tag{7}$$

[4]An odd function verifies $f(-x) = f(x)$. For the logistic function, $F(x) - 1/2$ is odd.
[5]An even transition function verifies $f(-x) = f(x)$.

was made popular by Terasvirta (1994) for modelling size asymmetries in models presenting a non-linearity in the mean. Here $F(\pm\infty) = 1$ and $F(0) = 0$. So α_1 will characterise small shocks and α_2 big shocks. For $\gamma \to \infty$, $F(.)$ becomes an indicator function for the point $u = 0$, which makes our model equivalent to the symmetric GARCH. Hagerud (1997) suggested to use this transition function for size asymmetry in GARCH models. However, this simple exponential function is rather restrictive as it does not give an information on the magnitude of what is really a big shock. An improvement over this function[6] that says that F goes to zero if u belongs to the interval $[-c, c]$ and goes to one otherwise is given by:

$$F(u_{t-1}, \gamma, c) = \frac{1 - \exp(-\gamma u_{t-1}^2)}{1 + \exp[-\gamma(u_{t-1}^2 - c^2)]} \tag{8}$$

I have now introduced an extra parameter, the threshold c, that determines at which magnitude of past errors the change of regime occurs. The range of c is restricted to positive values for an identification purpose.

The exponential transition function (7) can also be generalised by introducing a threshold parameter c

$$F(u_{t-1}, \gamma, c) = 1 - \exp[-\gamma\,(u_{t-1} - c)^2] \tag{9}$$

The parameter c translates the exponential curve so that $F(.)$ is now minimum at $u_t = c$. This introduces the possibility of an asymmetry between positive and negative shocks that is now combined with the size effect as will be detailed in the next section.

3 Comparing models through news impact curves

A clear idea about the differences between the three above proposed smooth transition models is provided by the inspection of the "news impact curve". As defined by Engle and Ng (1993), the news impact curve gives the relation between h_t and u_{t-1} holding constant the other informations. It shows how new information is incorporated to the measure of volatility. We shall see that the different models I have proposed have very different news impact curves. This means that predicted volatility, at least in the short term, will

[6]See the paper of Jansen and Terasvirta (1996) for another type of generalisation of the exponential transition function used for non-linear modelling of the mean.

be very different according to the model which is used. This has no trivial consequences on the valuation of options, on hedging strategies or portfolio selection.

The news impact curve of the STGARCH is

$$h = \omega + \alpha_1 u^2 + \lambda\, u^2\, F(u,\gamma) + \beta \bar{h} \tag{10}$$

where $\bar{h}$ is set equal to the unconditional mean of the returns. Fixing ω, α_1, γ and β and giving a value to $\bar{h}$, the function h can be drawn for the whole range of possible values for u. In the next subsection, I shall take a range for u equal to $[-4, 4]$. I shall fix the parameters at the following values: $\omega = 0.25$, $\alpha_1 = 0.3$, $\lambda = -0.25$, $\beta = 0.6$. The parameters γ and c will vary from case to case.

3.1 Logistic and exponential

Let me now draw various news impact curves for specified values of the parameters. The news impact curve of the simple GARCH is symmetric for

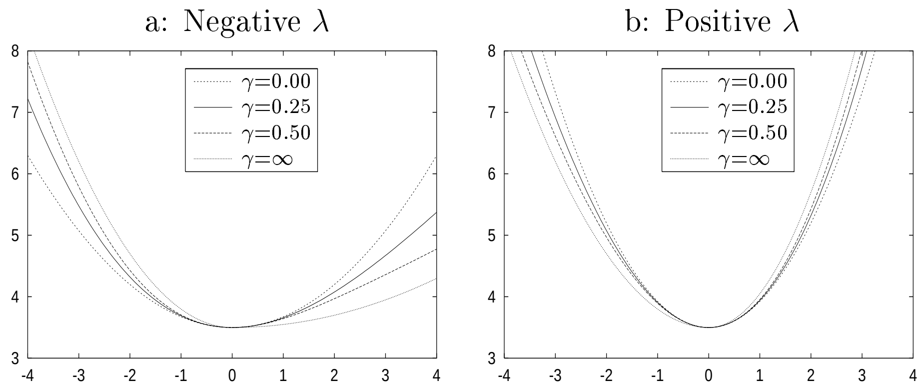

Figure 1: News impact curves for LSTGARCH

negative and positive news. The logistic transition function operates a rotation around $u = 0$ in the LSTGARCH, on the right for negative a λ (see Figure 1.a) or on the left for a positive λ (see Figure 1.b). The importance of the rotation is determined by γ but has a limit given by the TGARCH impact curve ($\gamma = \infty$). The impact curve of the EGARCH model has a similar general shape, but its tails are of course of the exponential form.

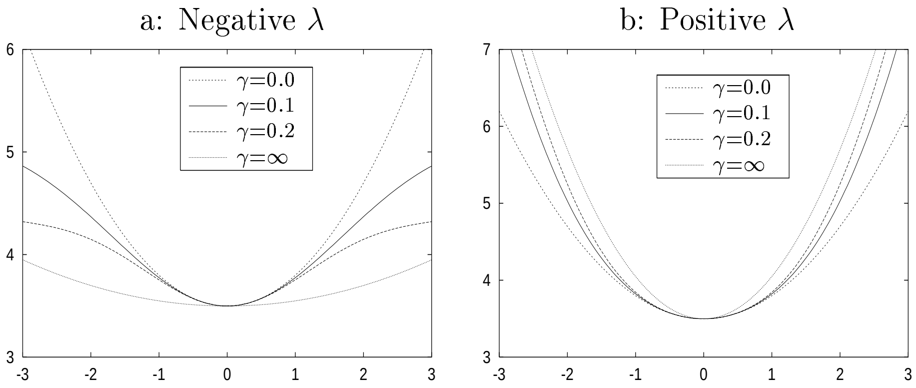

Figure 2: News impact curves for ESTGARCH

The exponential transition function (with $c = 0$) gives more weight to small news compared to the symmetric GARCH, and soften greatly the impact of big news in the ESTGARCH (see Figure 2.a). The point at which the impact curve crosses the impact curve of the simple GARCH depends on γ. But γ monitors also smoothness of the transition. The model is relatively constrained as for either $\gamma = 0$ or $\gamma \to \infty$, we recover a symmetric GARCH. A U shape can be given to the news impact curve by changing the sign of λ (see Figure 2.b). In that case, big news receive more weight than small ones.

3.2 The importance of the threshold parameter c

Introducing a threshold c gives a great flexibility in managing the impact curves of models with exponential and generalised exponential transition functions[7]. Let me consider first the generalised exponential model or GESTGARCH. The introduction of a positive c (or negative as the transition function is symmetric in c) together with a small value of γ modifies roughly speaking the level of the tails, but not their slope (see Figure 3.a). So the constraints present in the initial ESTGARCH are removed. A higher value for γ distorts the impact curve more deeply. Introducing a threshold in the exponential transition function mixes two asymmetry effects: the size and the sign effect. For small values of γ, a negative c rotates the impact curve clockwise and a positive c anticlockwise (see Figure 3.b). A higher value of γ gives rise to humps in the right or the left tail depending on the sign of

[7] A threshold, when introduced in the logistic transition function, has a very minor influence on the shape of the impact curve. Consequently it is not worth being considered here.

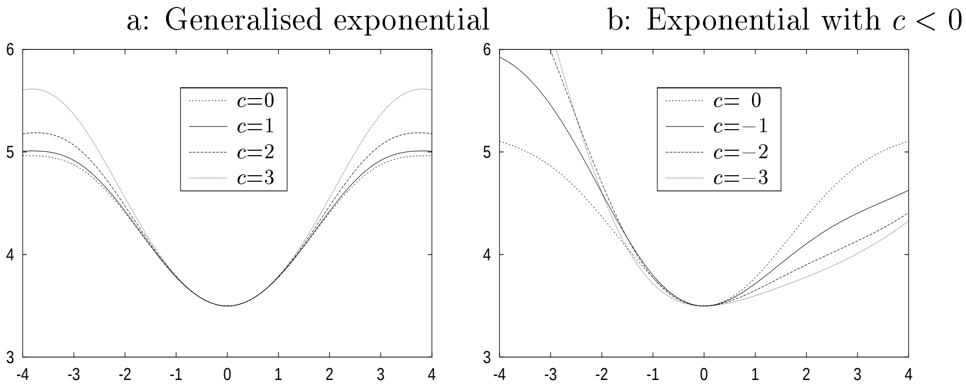

Figure 3: News impact curves for STGARCH with a threshold

c. A positive λ gives a U shape to the impact curve and a negative threshold translate and distorts the left tail, while leaving the right tail roughly unchanged.

4 Specification searches

The multiplicity of possible non-linear models that can be applied to a given sample renders the econometric analysis difficult. Tools are needed for a preliminary analysis especially here as the computational burden for inference may be important. We must first verify if a GARCH model is necessary and second if the GARCH effect presents a non-linearity of one of the types described above. Bauwens and Lubrano (1991) have proposed a Bayesian approach to investigate the Bayesian residuals of a linear regression model when heteroskedasticity is suspected. The method compares the posterior expectation of the squared Bayesian residuals computed under the null hypothesis of homoskedasticity H_0, with the predictive mean of the squared error term computed under various hypothesis of heteroskedasticity H_i. The posterior distribution of the Bayesian residuals contains information on what the residuals really are, including modelling deficiencies. On the contrary, the predictive density of the error term of a model predicts what the residuals should be considering a specific alternative hypothesis. This duality will be at the heart of my specification strategy where I shall compare posterior and predictive expectations by mean of an auxiliary regression using a Bayesian information criterion to select the most likely hypothesis.

4.1 Posterior residuals under H_0

In a linear regression model, a Bayesian residual (as defined for instance in Zellner 1975), is a random variable that is a linear combination of the random variable δ once a given sample (y, X) of dimension T has been observed:

$$u = y - X\delta \tag{11}$$

The variable u also represent the news as defined above. The posterior distribution of u is a linear transformation of the posterior distribution of δ. Under H_0 and a natural conjugate prior density, the latter is a Student density:

$$\varphi(\delta|y) = f_t(\delta|\delta_*, M_*, s_*, \nu_*) \tag{12}$$

where

$$\left\{ \begin{array}{lcl} M_* & = & M_0 + X'X \\ \delta_* & = & M_*^{-1}(M_0\delta_0 + X'y) \\ s_* & = & s_0 + y'y + \delta_0' M_0 \delta_0 - \delta_*' M_* \delta_* \\ \nu_* & = & \nu_0 + T - d \end{array} \right. \tag{13}$$

result from the application of standard formulae of natural conjugate analysis[8]. Symbols with a subscript equal to zero represent prior quantities and those with a subscript equal to a star their posterior equivalent. Usual theorems on the distribution of linear transformations of random variables gives the posterior density of Bayesian residuals which is Student:

$$\varphi(u|y) = f_t(u|\hat{u}, P^+, s_*^2, \nu_*) \tag{14}$$

where P^+ is the Moore-Penrose inverse of:

$$P = XM_*^{-1}X' \tag{15}$$

and $\hat{u} = y - X\delta_*$. Under a non informative prior, the posterior mean of the Bayesian residual is given by the classical residual. By marginalisation, I get:

$$\varphi(u_t|y) = f_t(u_t|\hat{u}_t, (P_{tt})^{-1}, s_*^2, \nu_*) \tag{16}$$

and

$$\mathrm{Var}(u_t|y) = \frac{\nu_*}{\nu_* - 2} s_*^2 P_{tt} \tag{17}$$

As $\mathrm{Var}(u_t|y) = \mathrm{E}(u_t^2|y) - [\mathrm{E}(u_t|y)]^2$, I can deduce that

$$\mathrm{E}(u_t^2|y) = \mathrm{E}(\sigma^2|y)\, P_{tt} + \hat{u}_t^2 \tag{18}$$

[8]See for instance Bauwens et al (1999), chapter 2

For $t \to \infty$, $\mathrm{E}(u_t^2|y) \to \hat{u}_t^2$ as P_{tt} represents a small sample effect that goes to zero. The posterior expectation $\mathrm{E}(u_t^2|y)$ will be compared to various predictive expectations by means of the following regression:

$$\hat{u}_t^2 = a_0 + a_1 P_{tt} + a_2 \sum_{j=1}^{p} \frac{p+1-j}{p(p+1)} g_{H_i}(\hat{u}_{t-j}^2) + \epsilon_t \tag{19}$$

The left hand side plus a_0 and $a_1 P_{tt}$ represents the posterior expectation of u_t^2 under H_0. The rest of the right hand side represents the predictive expectation of u_t^2 under alternative hypotheses of non-linearity. To limit the number of regressors which is the plague of this type of regression[9], I have used the approximation defined by Engle (1982) to enforce positivity constraints for ARCH(p). It is important to keep the number of regressors at a minimum to conduct the search for the type of non-linearity. Otherwise, the message of the sample may be blurred by too many regressors.

4.2 Predictive errors under H_i

The predictive of the observed sample under H_0 is

$$\begin{aligned} p(u_t) &= \int p(u_t|\sigma^2)\,\varphi(\sigma^2)\,d\sigma^2 \\ &= \int f_N(u_t|0,\sigma^2)\,f_{ig}(\sigma^2|s_0,\nu_0)\,d\sigma^2 \\ &= f_t(u_t|0,1,s_0,\nu_0) \end{aligned} \tag{20}$$

where the subscripts N, ig and t stand for the Normal, the inverted gamma2 and the Student densities. See Bauwens et al (1999), appendix A for more details. Under H_i, σ^2 becomes time variable so that the predictive density becomes

$$p(u_t) = \int p(u_t|h_t(\theta))\,\varphi(\theta)\,d\theta \tag{21}$$

But this density and its moments have no closed form. It is convenient for the sequel to consider an approximate predictive density with

$$\begin{aligned} p(u_t) &= \int p(u_t|h_t)\,\varphi(h_t)\,dh_t \\ &= \int f_N(u_t|0,h_t)\,f_{ig}(h_t|h_t^0,\nu_0)\,dh_t \\ &= f_t(u_t|0,1,h_t^0,\nu_0) \end{aligned} \tag{22}$$

where $f_{ig}(h_t|h_t^0,\nu_0)$ is a prior density on h_t with scale parameter h_t^0 and ν_0 degrees of freedom. A part of the model is for the while discarded, but will

[9]See for instance Granger and Terasvirta (1993), page 73 where a similar type of regression is used for testing linearity in the mean.

be reintroduced by an adequate choice of the scale parameter h_t^0. The approximate predictive density of the squared error term u_t^2 is a Fisher density:

$$u_t^2 \sim F(1, \nu_0)\, h_t^0/\nu_0 \tag{23}$$

with expectation:

$$\mathrm{E}(u_t^2) = h_t^0/(\nu_0 - 2) \tag{24}$$

I shall concentrate all my efforts in determining the most reasonable h_t^0 corresponding to each alternative hypothesis.

4.2.1 H_1: ARCH(1) model

In an ARCH(1) model, the skedastic function is $h_t = \omega + \alpha\, u_{t-1}^2$ with $u_t = y_t - x_t\delta$. I suppose that x_t is univariate for ease of notations. Combining these two expressions gives:

$$h_t = \omega + \alpha y_{t-1}^2 + \alpha\delta^2 x_{t-1} - 2\alpha\delta y_{t-1} x_{t-1} \tag{25}$$

Let me linearise the products $\alpha\delta$ and $\alpha\delta^2$ around $\alpha = 0$ and $\delta = \hat{\delta}$ which is the OLS estimator of δ under H_0. The resulting candidate for h_t^0 is:

$$h_t^0 = \omega + \alpha \hat{u}_{t-1}^2 \tag{26}$$

4.2.2 H_2: STARCH(1) model

Combining (5) and (26), the scale parameter of my prior is now of the form:

$$h_t^0 = \omega + \alpha_1 \hat{u}_{t-1}^2 + \lambda\, \hat{u}_{t-1}^2 F(u_{t-1}, \gamma) \tag{27}$$

This formulation is not very convenient for the while because it still presents the non-linearity. I shall apply the same recipe as above and linearise $F(u_{t-1}, \gamma)$. This is in accordance with some of the linearity tests reported in the classical literature [see e.g. Luukkonen et al (1988) and Terasvirta (1994)] where a third order Taylor expansion of $F(z)$ around $z = 0$ is recommended. The motivation in the classical literature is to overcome the identification problem present under the null of linearity. Here the question is simply to get a manageable prior scale parameter to compute an approximate conditional predictive density. I shall select the logistic, exponential and generalised exponential transition functions. Their respective third order Taylor expansions are:

$$\begin{array}{lll} T_{3L}(z) & \propto & 24 + 12z - z^3 \\ T_{3E}(z - c) & \propto & 3c^2 - 6cz(1 - c^2) + 3z^2(1 - 3c^2) + 2cz^3(3 - 5c^2) \\ T_{3GE}(z^2 - c^2) & \propto & z^2(2 - c^2) \end{array} \tag{28}$$

Going up to the third order makes very apparent the difference between these three transition functions. The expansion of the logistic is always of the first and third degree, while the expansion of the exponential with $c = 0$ and of the generalised exponential are of the second degree only. The exponential with a threshold has an expansion where all the terms up to the third degree are present. Of course, for $c = 0$, the adequate term vanish.

Let me now replace $F(u_{t-1}, \gamma)$ by its Taylor expansion in the expression of h_t^0 (27). I get three alternative formulations for the prior scale parameter

$$h_t^0 = \omega + \alpha_1 \hat{u}_{t-1}^2 + \lambda \hat{u}_{t-1}^2 \times \begin{cases} (c_0 + c_1 \hat{u}_{t-1} - c_3 \hat{u}_{t-1}^3) \\ (c_0 - c_1 \hat{u}_{t-1} + c_2 \hat{u}_{t-1}^2 + c_3 \hat{u}_{t-1}^3) \\ (c_0 + c_2 \hat{u}_{t-1}^2) \end{cases} \tag{29}$$

which will be the basis of my specification search.

4.3 Comparing hypotheses using an auxiliary model

I can now compare the posterior expectation of u_t^2 to its various possible predictive expectations by mean of the following general regression:

$$\begin{aligned} \hat{u}_t^2 \quad = \quad & a_0 + a_1 P_{tt} + a_2 \textstyle\sum_{j=1}^p \dfrac{p+1-j}{p(p+1)} \hat{u}_{t-j}^2 \times \\ & [1 + a_3 \hat{u}_{t-j} + a_4 \hat{u}_{t-j}^2 + a_5 \hat{u}_{t-j}^3] + \epsilon_t \end{aligned} \tag{30}$$

I first have to select the optimal p in the full regression (30) using an information criterion. Conditionally on that choice, I shall try to impose the restrictions corresponding to each of the three possible models and choose the model that minimises the Schwarz criterion . This corresponds to selecting the model that has the maximum posterior probability under a diffuse prior information. The set of restrictions can be summarised as follows:

- $a_2 = a_3 = a_4 = a_5 = 0$ mean that there is no ARCH effect.

- $a_3 = a_4 = a_5 = 0$ mean that the ARCH effect is linear.

- $a_4 = 0$ while $a_3 \neq 0$ and $a_5 \neq 0$ means that the non-linearity is of an odd type corresponding to a logistic transition function.

- $a_3 = a_5 = 0$ and $a_4 \neq 0$ means that the non-linearity is of an even type that can be modelled either by an exponential with $c = 0$ or a generalised exponential transition function.

- $a_3 \neq 0$, $a_4 \neq 0$ and $a_5 \neq 0$ means that the non-linearity mixes size and sign effects which can be reasonably modelled by an exponential transition function with a threshold, although this case is not as clear as the others. Many non-linear transition functions can lead to such a Taylor expansion.

Remark:

Engle and Ng (1993) have proposed in a classical framework sign and size bias test regressions as mis-specification tests, but also as exploration tools. They regress $\hat{u}_t^2$ over a constant and S_{t-1} for positive sign bias, $S_{t-1}\hat{u}_{t-1}$ for positive size bias and $(1 - S_{t-1})\hat{u}_{t-1}$ for negative size bias.

5 Bayesian inference in the STARCH model

Bayesian inference in usual GARCH models is made complicated by the fact that the posterior density must be integrated out numerically. Kleibergen and van Dijk (1993) use importance sampling while Geweke (1994) proposes an independent Metropolis algorithm. The usual Gibbs sampler is not applicable here as the posterior density has no property of conjugacy [as defined by Carlin and Gelfand (1991)]. Bauwens and Lubrano (1998) use a Griddy-Gibbs sampler to overcome this difficulty. The algorithm is based on the numerical inversion of each conditional posterior density. As in usual Gibbs samplers, the algorithm encounters numerical difficulties in the case of a strong correlation between the parameters of the model. This may motivate the introduction of an alternative parameterisation of the GARCH model. The presence of a smooth transition function creates some specific problems that concerns the integrability of the posterior density of γ, the smooth transition parameter and a local identification problem at $\gamma = 0$ that corresponds to the linearity of the skedastic function h_t. The threshold c may also cause identification problems.

5.1 Prior and posterior densities

Let me give now the complete notation of the smooth transition GARCH model:

$$\begin{cases} y_t = x_t'\delta + u_t \\ u_t = \epsilon_t\sqrt{h_t} \qquad \epsilon_t \sim N(0,1) \\ h_t = \omega + \alpha_1 u_{t-1}^2 + \lambda u_{t-1}^2 F(u_{t-1}, \gamma, c) + \beta h_{t-1} \end{cases} \tag{31}$$

Let me define the diagonal $(T-1)\times(T-1)$ matrix $H(\delta,\theta)$ having h_t as its $[t,t]$ element and where $\theta' = [\omega, \alpha_1, \lambda, \beta, \gamma, c]$. The likelihood function of the $T-1$ observations of y_t is:

$$l(y;\delta,\theta) \propto |H(\delta,\theta)|^{-1/2} \exp -\frac{1}{2} u' H^{-1}(\delta,\theta) u \tag{32}$$

where $u = y - X\delta$. To evaluate this function, it is necessary to define h_1. I shall treat this initial condition as fixed. For $u_0 = 0$, $h_1 = \omega$. So h_1 can be taken equal to the empirical variance of the first observations.

I shall be non informative on δ with

$$\varphi(\delta) \propto 1 \tag{33}$$

It is difficult to devise an informative prior for the skedastic parameters. The prior should guaranty or preserve the positivity of h_t. In a usual GARCH, this means that ω, α and β should be positive. Here α_1 and α_2 have to be positive, which means that $\lambda > -\alpha_1$. So the parameterisation in λ may not be the most convenient to impose positivity constraints a priori. In the transition function $F(.)$, γ has to be positive for identification purposes. The same may hold for c in certain transition functions. I shall not try to impose a priori strict stationarity that is necessary for inference. But in general, strict stationarity is verified for most available samples [see e.g. Kleibergen and van Dijk (1993)].

I first propose a non informative prior for the parameters of the skedastic function that cause no problem. I shall discuss below the case of $\varphi(\gamma)$ and of $\varphi(c)$ that raise specific issues such as integrability of the posterior and identification. So:

$$\varphi(\omega, \alpha_1, \alpha_2, \beta) \propto \begin{cases} 1 & \text{if } \omega, \alpha_1, \alpha_2, \beta > 0 \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

The complete posterior density

$$\varphi(\delta,\theta|y) \propto \varphi(\delta) \times \varphi(\omega, \alpha_1, \alpha_2, \beta) \times \varphi(\gamma, c) \times l(y;\delta,\theta) \tag{35}$$

has to be integrated out numerically. No partial analytical integration is possible.

5.2 Integrability for γ

The smooth transition function $F(.)$ becomes a Dirac function for $\gamma \to \infty$. For instance with the logistic function model (31) becomes in this case observationally equivalent to the model of GJR (1993). Consequently:

Theorem 1 *The posterior density of γ is of the same order of integrability as the prior $\varphi(\gamma)$. In particular posterior moments exist only up to the order of prior moments.*

Proof: For $\gamma \to \infty$, $F(\gamma, u_t)$ is $O(1)$ in γ. Consequently $\varphi(\gamma|y)$ does not tend to 0 as γ tends to infinity and is not integrable.

□

A similar problem arises in regression models with Student errors as underlined in Bauwens and Lubrano (1998). The conclusion is that a 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 $O(\gamma^{1+v})$ with $v > 0$. A convenient minimal prior is the truncated Cauchy density with [10]:

$$\varphi(\gamma) = \begin{cases} (1 + (\gamma - \gamma_0)^2)^{-1} & \text{if } \gamma > 0 \\ 0 & \text{otherwise} \end{cases} \tag{36}$$

Its right tail is $O(\gamma^2)$. This prior has no moment. Only the mode and the quantiles exist[11]. The prior mode of γ is equal to γ_0. In order to be able to elicit easily γ_0, γ should be scale free. This is obtained by scaling the observations (dividing the y_t by their empirical standard deviation). With $\gamma_0 = 0$, we have the least informative case. Increasing γ_0 leads to a model where the transition function is sharper and sharper, leading at the limit to model (4) if the transition function is odd. I shall perform a sensitivity analysis with varying γ_0, while being non-informative on the other parameters.

5.3 Local identification for λ

The smooth transition function becomes constant at the point $\gamma = 0$. Consequently the skedastic function presents a perfect collinearity and the parameter λ becomes not identified. If for some samples, the point $\gamma = 0$ is not in the useful integration range and so the problem has no practical importance, this is not the general situation, especially for high frequency data.

[10] 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 $\gamma > 0$.

[11] A prior that guaranties the existence of all the posterior moments of γ is the exponential prior used for instance by Geweke (1993):

$$\varphi(\gamma) \propto \exp(-\gamma/\gamma_0) \qquad \gamma_0, \gamma > 0 \tag{37}$$

Consequently it is recommended to exclude a priori the point $\gamma = 0$ from the integration range. It may also be wise to use a positive γ_0 in the Cauchy prior, so as to eventually translate the mode of the posterior away from zero. Supposing γ strictly positive is not a bias in favour of asymmetry as there is still the possibility that $\lambda = 0$.

5.4 The special case of the threshold parameter c

In usual non-linear regression models, the posterior density of the threshold c may be very badly behaved [see for instance Lubrano (1998) or Osiewalski and Welfe (1998)]. Here, the threshold parameter determines what is the most likely value for a big shock when the generalised exponential transition function is chosen. When a threshold is used in the exponential transition function, it determines where, on the scale of the u is the asymmetry between the positive and negative shocks. Of course the domain of definition of c is determined by the observed sample. But there will be an identification problem every time there is not enough observations left between the end points of the integration range and the minimum and maximum of the sample. So c has to be integrated out on a restricted range.

5.5 Another possible parameterisation

The general notation for the GARCH model assumes that in (1) the variance of ϵ_t is not constrained to be one, but is equal to σ^2. Adopting this formulation for the STGARCH model gives:

$$\begin{cases} y_t = x_t'\delta + u_t \\ u_t = \epsilon_t \sqrt{h_t} \qquad \epsilon_t \sim N(0, \sigma^2) \\ \sigma^2 h_t = \omega + \alpha_1 u_{t-1}^2 + \lambda\, u_{t-1}^2 F(u_{t-1}, \gamma, c) + \beta \sigma^2 h_{t-1} \end{cases} \tag{38}$$

I can divide by σ^2 both members of the definition of h_t:

$$h_t = \frac{\omega}{\sigma^2} + \frac{\alpha_1}{\sigma^2} u_{t-1}^2 + \frac{\lambda}{\sigma^2} u_{t-1}^2 F(u_{t-1}, \gamma, c) + \beta h_{t-1} \tag{39}$$

This shows that in this parameterisation only $\tilde{\omega} = \omega/\sigma^2$, $\tilde{\alpha}_1 = \alpha_1/\sigma^2$ and $\tilde{\lambda} = (\alpha_2 - \alpha_1)/\sigma^2$ are identified. On the contrary, β is always identified. The traditional parameterisation insures identification by setting $\sigma^2 = 1$. Another possible simple identification rule consists in setting $\tilde{\omega} = 1$ which means that $\omega = \sigma^2$. The drawback of this parameterisation is that $\tilde{\alpha}_1$ and $\tilde{\lambda}$ now depend on the scale of the y_t. Its advantage is that σ^2 can be integrated out analytically. The size of the numerical integration is thus reduced by one. The skedastic function becomes:

$$h_t = 1 + \tilde{\alpha}_1 u_{t-1}^2 + \tilde{\lambda} u_{t-1}^2 F(u_{t-1}, \gamma, c) + \beta h_{t-1} \tag{40}$$

As σ^2 will be integrated out analytically, there will be no possible numerical trade off between the constant and the variable part of the skedastic function. This certainly will improve numerical stability [see e.g. Robert and Mengersen (1995) for an analysis of reparameterisation issues on the performance of the Gibbs sampler].

The likelihood function (32) of the $T-1$ observations of y_t is transformed into:

$$l(y; \delta, \sigma^2, \tilde{\theta}) \propto \sigma^{-(T-1)} |H(\delta, \tilde{\theta})|^{-1/2} \exp -\frac{1}{2\,\sigma^2} u' H^{-1}(\delta, \tilde{\theta}) u \tag{41}$$

where $\tilde{\theta}' = [\tilde{\alpha}_1, \tilde{\lambda}, \beta, \gamma, c]$. Note that in order to evaluate this likelihood function, I have now $h_1 = 1$ as a starting value.

The prior densities on σ^2 and $\tilde{\theta}$ have to be made compatible with those on θ. A usual prior density on σ^2 which is a scale parameter is

$$\varphi(\sigma^2) = f_{ig}(\sigma^2 | s_0, \nu_0) \tag{42}$$

or simply $\varphi(\sigma^2) \propto \sigma^{-2}$ if we are non informative. As the new parameterisation implies that $\omega = \sigma^2$, the prior on ω must have the same form. The prior density (34) on α and λ was uniform. Due to the Jacobian of the transformation from α_i to $\tilde{\alpha}_i$, the resulting prior is

$$\varphi(\alpha_1, \lambda) \propto \sigma^{-4} \tag{43}$$

The priors on β, γ and c remain the same as before. With these priors, I can get nice expressions for the posterior densities

Theorem 2 *The marginal posterior density of δ and $\tilde{\theta}$ is given by*

$$\begin{aligned}
\varphi(\delta, \tilde{\theta} | y) &\propto \int_{\sigma^2 > 0} l(y; \delta, \sigma^2, \tilde{\theta})\, \varphi(\sigma^2)\, \varphi(\delta)\, \varphi(\tilde{\theta})\, d\sigma^2 \\
&\propto |H(\delta, \tilde{\theta})|^{-1/2} \times [s_0 + u' H^{-1}(\delta, \tilde{\theta}) u]^{-(\nu_0 + T + 3)/2}
\end{aligned}$$

The conditional posterior density of σ^2 is an inverted gamma2 with:

$$\varphi(\sigma^2 | \delta, \tilde{\theta}, y) = f_{ig}(\sigma^2 | s_*(\delta, \tilde{\theta}), \nu_0 + T + 3) \tag{44}$$

where

$$s_*(\delta, \tilde{\theta}) = s_0 + u' H^{-1}(\delta, \tilde{\theta}) u \tag{45}$$

The marginal posterior density of σ^2 can be simulated by:

$$\sigma_j^2 \sim s_*(\delta_j, \tilde{\theta}_j)/\chi^2(\nu_0 + T - 1) \tag{46}$$

where δ_j and $\tilde{\theta}_j$ are draws of the corresponding parameters made from their marginal posterior density.

Proof: The proof follows from the integrating constants of the inverted gamma2 density. The simulation procedure is simply based on the property of usual algorithms for simulating an inverted gamma2.

□

Of course, we are interested in the posterior density of the original parameterisation which is the most common one and not of the transformed parameters. The backward transformation can be done easily. Posterior draws for α_1 and λ are obtained using the non-linear transformation: $\alpha_{1j} = \tilde{\alpha}_{1j} \times \sigma_j^2$ and $\lambda_j = \tilde{\lambda}_j \times \sigma_j^2$.

6 Specification and inference for the Brussels index

As a first illustration of the method, I shall use data of the Brussels Spot Market Index. I have weekly data collected at the closure of the market each Friday covering the period 03/01/86 to 26/01/96. The data have been corrected for bank holidays. This make 509 observations that I will take in differences of logs. The transformed observations are displayed in Figure 4. These data present very large fluctuations after the melt down of October 1987, some large fluctuations around 1991 and are relatively quiet after 1993.

6.1 Specification search

The returns are first filtered by an AR(2) model to remove auto-correlations:

$$\Delta y_t = \underset{[1.74]}{0.08} + \underset{[4.41]}{0.19}\Delta y_{t-1} + \underset{[3.68]}{0.16}\Delta y_{t-2} + \hat{u}_t$$

(t statistics are given between square brackets). From this regression, I get 506 residuals $\hat{u}_t$ that are analysed by mean of the auxiliary regression (30). A BIC determined an optimal p of 4. The resulting regression is:

$$\begin{aligned}\hat{u}_t^2 \;=\;& \underset{[1.10]}{0.15} + \underset{[3.10]}{0.00025}P_{tt} + \underset{[2.56]}{0.50}[\textstyle\sum_{j=1}^4 \dfrac{p+1-j}{p(p+1)}\hat{u}_{t-j}^2]\times \\ & [1 - \underset{[-8.42]}{0.51}\,\hat{u}_{t-j} - \underset{[-1.25]}{0.011}\,\hat{u}_{t-j}^2 + \underset{[8.09]}{0.019}\hat{u}_{t-j}^3] + e_t\end{aligned}$$

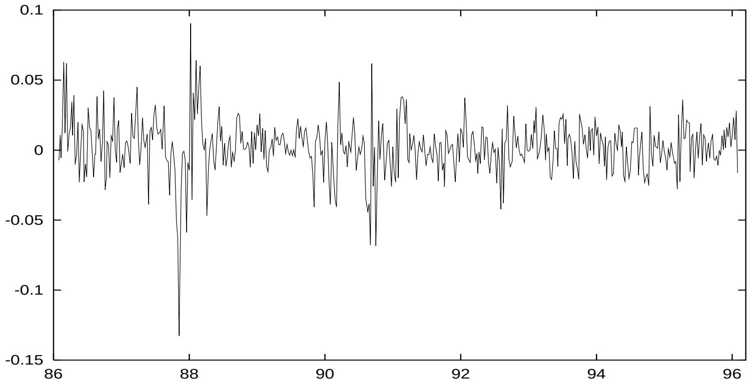

Figure 4: Brussels Spot Index returns

Table 1: Which GARCH for SPOTW data

	linear	non-linear	odd transition	even trans.
BIC	1.729	1.567	1.555	1.672

Let me now investigate various types of restrictions in this regression and select the most probable model, i.e. the one that has the minimum Schwarz criterion. From Table 1, I can conclude that there is definitely an ARCH effect, that this effect is non-linear and that the non-linearity distinguishes between negative and positive shocks. The non-linearity is generated by the large fluctuations of the first half of the sample. If I discard the first 250 observations, the ARCH effect becomes linear. But if I discard only the first 100 observations non-linearity is still present, despite the fact that the melt down of 1987 is excluded. For each sub sample analysis, I refiltered the data with an updated AR(2).

6.2 Inference results

A LSTGARCH(1,1) model was fitted to the filtered returns that were standardised (division by the empirical standard deviation). Maximum likelihood estimates, as reported in Table 2, were obtained with a standard algorithm. Their Bayesian counterpart[12] corresponds to the case were $\gamma_0 = 0$, the prior

[12]The draws of the Griddy Gibbs display a certain amount of negative correlation between β and $(\omega,\ \alpha_1)$ of respectively -0.79 and -0.63. Other correlations are negligible. I used 1000 draws + 150 draws for warming up the chain Convergence was checked using

on all the other parameters being uniform. There are not much differ-

Table 2: MLE and posterior results
for SPOTW data

	ω	α_1	α_2	β	γ
MLE	0.055 [0.020]	0.24 [0.061]	0.051 [0.042]	0.79 [0.042]	1.84 [2.60]
$\gamma_0 = 0$	0.069 [0.026]	0.29 [0.072]	0.064 [0.042]	0.75 [0.051]	1.91 [1.60]
$\gamma_0 = 10$	0.070 [0.024]	0.26 [0.056]	0.079 [0.039]	0.76 [0.044]	9.89 [2.36]
$\gamma_0 = 100$	0.073 [0.028]	0.26 [0.063]	0.081 [0.042]	0.75 [0.055]	100 [3.57]
$\alpha_1 = \alpha_2$	0.075 [0.028]	0.19 [0.044]	-	0.73 [0.053]	-

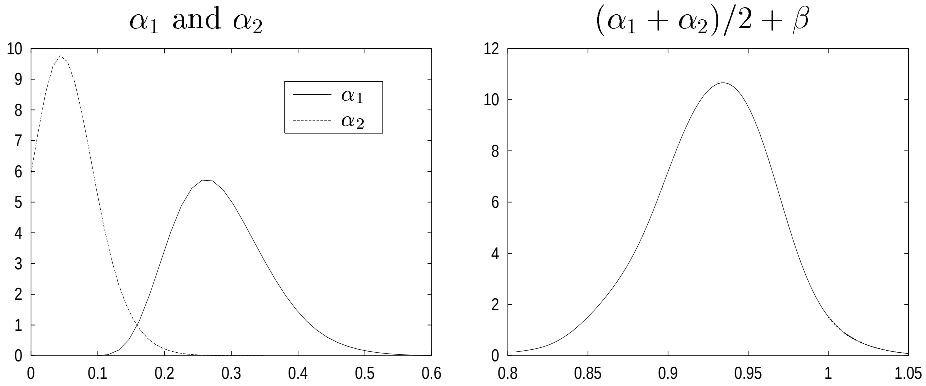

Figure 5: Posterior densities for BRUSSELS
with $\gamma_0 = 0$

ences between classical and Bayesian estimates in this case. On average, Bayesian standard deviations are slightly greater, but the ratio of posterior means over posterior standard deviations are in general lower. Note that the smooth transition parameter γ is, as usual, badly estimated. Increasing the value of γ_0 takes us nearer to the GJR model. A truly abrupt transition model ($\gamma_0 = 100$) is not confirmed by the data as the posterior expectation of ω, which represent the unexplained part of the volatility, slightly increases for $\gamma_0 = 100$. A symmetric GARCH model is also not favoured by the data,

CUMSUM graphs. It took 4 minutes on a Penthium 350 for the 1150 draws.

following the same criterion. Note that the posterior expectation of γ is very sensitive to the prior, showing that in fact the data may not be too informative on the speed of the transition.

The joint graph of the posterior densities of α_1 and α_2 shows in Figure 5 that negative (α_1) and positive (α_2) shocks have a quite different impact. Negative shocks have a greater contribution to the conditional variance than positive ones. If the graphs of the posteriors slightly overlap, the posterior probability that $\lambda > 0$ is equal to zero when estimated directly from the draws. The posterior graphs are obtained, not from the draws, but using a technique of variance reduction explained in Bauwens and Lubrano (1998).

The shocks are not persistent as the weak stationarity condition (see the appendix for a derivation) is verified with a probability of 0.98. The posterior density of the transformation $(\alpha_1+\alpha_2)/2+\beta$, given in Figure 5, was estimated with a kernel method, using the draws of the Gibbs sampler.

6.3 Analysis of non-linearity

The posterior density of γ for $\gamma_0 = 0$is given in Figure 6. It is fairly concentrated on low values of γ, which put credit to a smooth transition between the regimes. The moments of γ were computed on the truncated range [0,8],

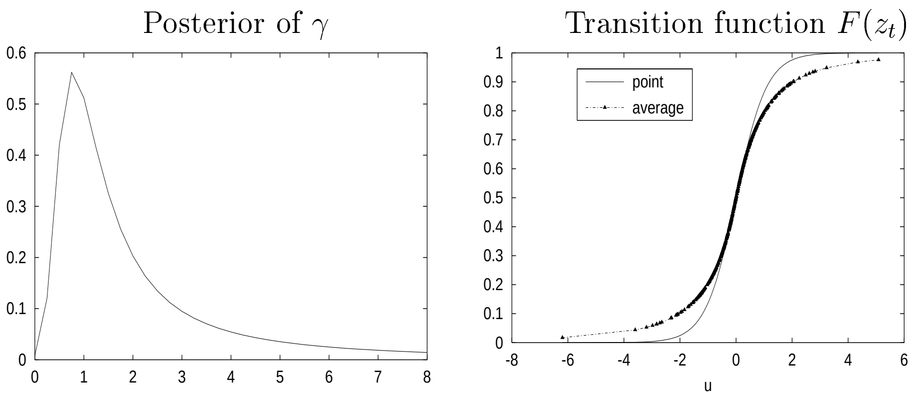

Figure 6: The smooth transition function for SPOTW data

because un-truncated posterior moments do not exist with a Cauchy prior. The graph of the smooth transition function $F(u_{t-1})$, given in Figure 6, is obtained as a transformation of the posterior draws. On the same graph, I give the same transition function, but evaluated at the posterior expectation.

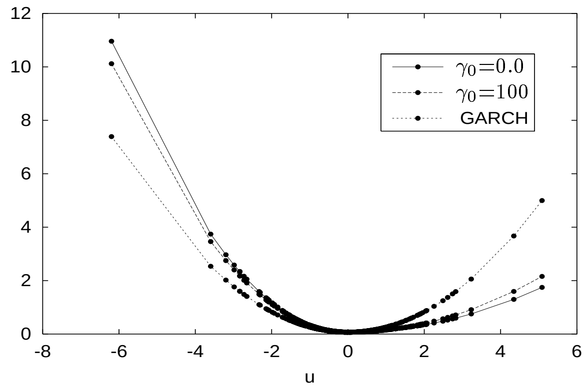

Figure 7: News impact curve for SPOTW data

This combination shows that the Bayesian transition function is much softer than the transition function obtained by a point estimate. As the filtered returns were standardised, the scale is expressed in term of standard deviations. Most of the observations are situated within ± 2 standard deviations. They correspond to the quasi linear part of the smooth transition function. The transition function is lower than 0.053 (respectively 0.10) or greater than 0.94 (respectively 0.90) only for four observations (respectively 10 and 8 observations which corresponds to the melt down of October 1987 and a period in July 1990. This demonstrates the fact that the smooth transition is important and that the change of regime is not an abrupt function of the signs of the shocks, but evolves gradually with their magnitude. This is a soft transition. An abrupt transition concerns only very few observations in the sample.

The news impact curve is displayed in Figure 7 for $\gamma_0 = 0$, for $\gamma_0 = 100$. This curve is very much tilted, revealing a large asymmetry between negative and positive shocks, even for small shocks. There is a visible difference between the smooth transition model and the GJR model. For ease of comparison, the news impact curve for the symmetric GARCH is also given. It underline the importance of modelling non-linearity.

7 Specification and inference for the Tokyo index

Here I try to illustrate the even transition function with the daily NIKKEI 225 index covering the period 17 Feb 1994 to 18 Jan 1996. This makes 475 observations corresponding to the closure of the market. The data, displayed in Figure 8, are taken in differences of logarithms. A first look at this graph

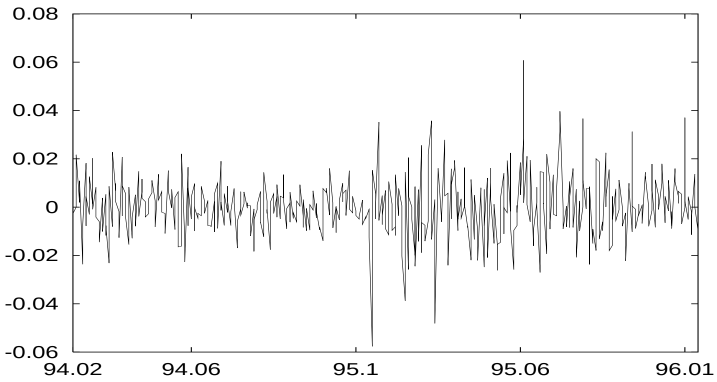

Figure 8: Tokyo NIKKEI 225 returns

reveals that we have another type of sample configuration. The first year of observations is characterised by a relative quietness of the market. The second year corresponds to a greater volatility with many peaks.

7.1 Specification search

It appeared that it was not necessary to filter the data. I only standardised the returns [13] $u_t = \Delta \log(y_t)$ where y_t is the level of the index. The test regression was estimated with an optimal lag of $p = 1$:

$$\begin{aligned} \hat{u}_t^2 &= \underset{[5.78]}{0.72} + \underset{[3.84]}{0.38}\,\hat{u}_{t-1}^2 \times \\ &\quad [1 + \underset{[0.071]}{0.0025}\,\hat{u}_{t-1} - \underset{[-3.02]}{0.016}\,\hat{u}_{t-1}^2 - \underset{[-0.10]}{0.00017}\,\hat{u}_{t-1}^3] + e_t \end{aligned}$$

(t statistics between square brackets). A sequence of BIC, as reported in Table 3 clearly indicates the even (exponential) transition model (ESTGARCH) as the preferred model. In preliminary inference results not reported here, I have tried three different even transition functions. The model with an exponential transition and a threshold seemed to give the better fit and the best account of non-linearity. The next subsection presents inference results corresponding to this case.

[13] This scaling was not enough for the Bayesian computations. It was necessary to divide the scaled returns by 10.

Table 3: Which GARCH for the daily NIKKEI data

	linear	non-linear	odd	even
BIC	1.612	1.635	1.640	1.605

7.2 Inference results

The Tokyo data set is tricky to analyse as there is a very high negative correlation between ω and β (around -0.95) that impeded the convergence of the Griddy Gibbs algorithm. However using the alternative parameterisation of the GARCH solves the case as it kills the correlation. With 1000+150 draws, convergence of the chain was achieved as checked on CUMSUM graphs. Thus the maximum likelihood estimates were computed using the normalisation $\sigma^2 = 1$ (and imposing positivity constraints). The Bayesian estimates were computed using $\omega = 1$. The Bayesian estimates reported in Table 4 were obtained by applying the necessary transformations indicated in Theorem 2 and below. I performed a sensitivity analysis and present results obtained with $\gamma_0 = 0$ (to be compared to the maximum likelihood estimates), with $\gamma_0 = 10$ and finally for a symmetric GARCH.

Table 4: Posterior results for the NIKKEI 250 index

	ω	α_1	α_2	β	γ	c
MLE	0.000020	0.59	0.069	0.87	6.28	-0.82
	[0.0025]	[0.17]	[0.030]	[0.048]	[5.21]	[0.18]
$\gamma_0 = 0$	0.0052	0.86	0.20	0.31	2.23	-1.04
	[0.0018]	[0.34]	[0.074]	[0.18]	[1.72]	[0.33]
$\gamma_0 = 10$	0.0051	1.27	0.22	0.32	10.4	-1.03
	[0.0019]	[0.54]	[0.077]	[0.18]	[2.56]	[0.18]
$\alpha_1 = \alpha_2$	0.0065	-	0.29	0.23	-	-
	[0.002]		[0.093]	[0.18]		

There are some substantial differences between the Bayesian and the classical estimates for the GARCH parameters and reasonable similarities for the non-linear parameters. The posterior density of β is in fact bimodal and this explains certainly the main differences. A Dickey-Savage ratio would favour the restriction $\beta = 0$, restriction which is in accordance with the specification search that retained an optimal lag $p = 1$ for the test regression. The probability that $\lambda = \alpha_2 - \alpha_1 > 0$ is equal to 0.015 (when $\gamma_0 = 0$ and 0.011 when $\gamma_0 = 10$). It goes up to 0.31 when c is imposed to be zero. The presence of a threshold is essential here for modelling the non-linear effect that totaly disappears otherwise. And the symmetric GARCH seems to be rejected by the data as the posterior expectation of ω is larger in this case as shown in the last line of Table 4. It is very difficult to say that big shocks have a

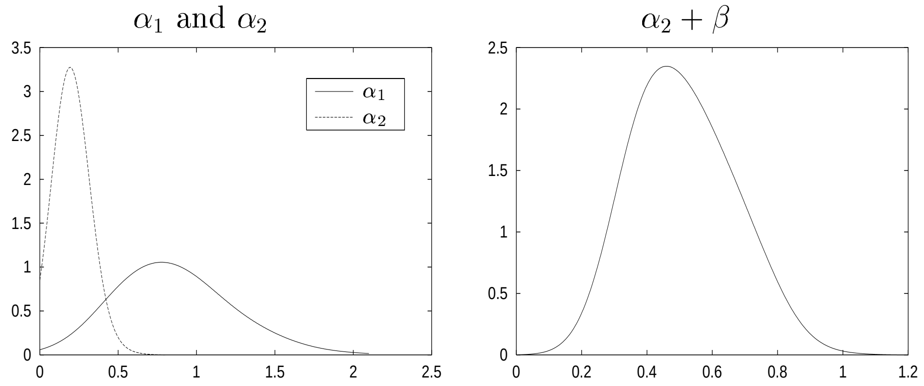

Figure 9: Posterior densities for Tokyo
with $\gamma_0 = 0$

permanent effect. The posterior expectation of $ws = \alpha_2 + \beta$, which gives indications on stationarity and persistence (see the appendix for a derivation), is 0.51 with a standard deviation of 0.14. The probability that $ws > 1$ is equal to 0.001. This probability does not change much when $c = 0$. The impact of bigger shocks given by α_2 is much smaller than that of smaller shocks (α_1). However, the posterior density of α_1 is flatter, indicating a larger uncertainty about the effective impact of small shocks on volatility.

7.3 Analysis of non-linearity

The even transition function is plotted against the standardised shocks. Most of the shocks are within ± 2 standard deviations. The transition function is sharper when evaluated at posterior expectation than when computed directly. Consequently the "Bayesian" actual transition is smoother than its "classical" counterpart. Averaging has also the consequence that the transition function is not zero at its minimum, except when $c = 0$. The presence of a threshold makes that there are much more observations on the right hand side of the transition function than on the left hand side even if there are as many positive than negative shocks in the data.

The graph of the posterior density of c shows that the negative threshold $c = -1.05$ is relatively precisely estimated. The range of integration for c was

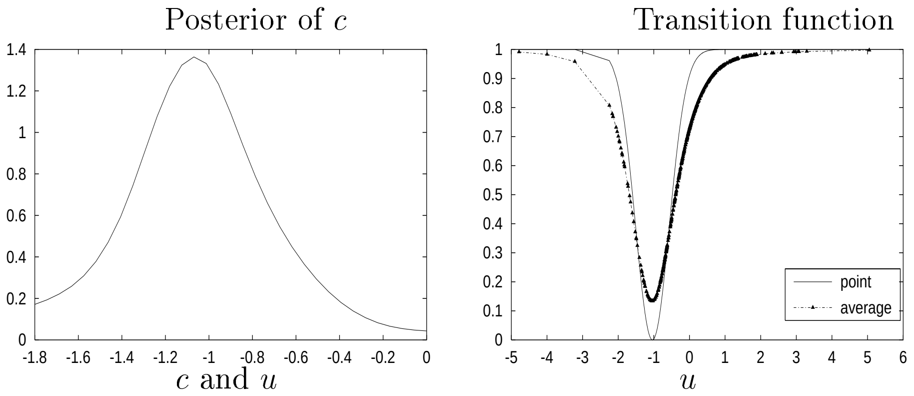

Figure 10: The smooth transition function for for Tokyo

truncated for identification reasons (greater than -1.8). The positive segment receives none of the probability [14] (as verified on experiments not reported here). When γ_0 is increased to 10, the posterior density of c concentrates around its posterior mean. The (Bayesian) news impact curve displays a

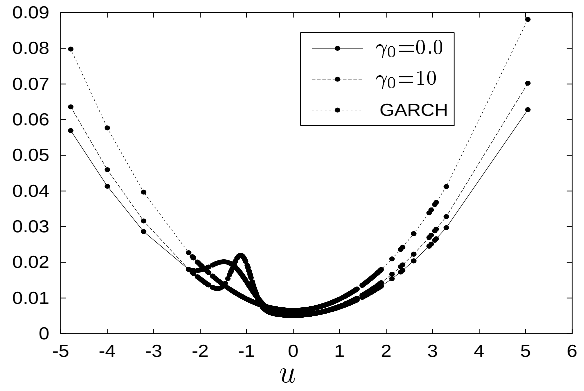

Figure 11: News impact curve for Tokyo

fairly large asymmetry when compared to the symmetric GARCH. Figure 11 indicates that if most of the points concern small shocks between ± 2 standard deviations, the deformation of the impact curve is situated inside that interval. Inside that interval, positive and negative shocks have a different impact. The volatility is higher for small negative shocks (between 0 and 1.5

[14] As a matter of fact, the maximum likelihood routine did not converge when c was left free, when c was constrained to be positive, but did converge when c was constrained to be negative.

standard deviations) than for positive ones. Over 2 standard deviations, positive and negative shocks have the same influence on volatility. The impact curve for the linear GARCH gives a greater weight to big shocks. Increasing γ_0 to 10 modified only locally the news impact curve.

8 Conclusion

In this paper, I have considered a new type of asymmetric GARCH(1,1) with different choices for the asymmetry. It appears clearly that the Bayesian approach was successful. First of all, the advantage of the Bayesian approach is that problems are clearly pointed out and in particular the difficulties coming from inference on the smooth transition parameter that are frequently reported in classical analysis of non-linear models [see e.g. Granger and Terasvirta (1993), p.123] are given here a sound theoretical basis. Secondly those inference difficulties receive a correct numerical treatment as the posterior density of the STGARCH model may reveal easier to integrate than to maximise in certain cases. Point estimates are not the same with the two methods because posterior densities are skewed and may exhibit bimodality.

The asymmetric GARCH models proposed by GJR (1993) or by Zakoian (1994) had an abrupt transition. With financial data, and especially at high frequency, the smooth transition proved to be relevant in the two empirical illustrations given in this paper. And the Bayesian approach produces a transition that is even smoother than the transition obtained with point estimates.

The presence or absence of non-linearity may depend on the frequency of the observations. In empirical investigations not reported in paper, I analysed the CAC40 Paris index. Weekly data did not present non-linearity according to the specification tests, but daily data did. This exercice demonstrates the usefulness of the specification tools introduced in this paper to detect the presence or absence of non-linearity in the data and their ability to give some information on the nature of non-linearity. The question still remains to know if the models I have proposed and implemented capture all the asymmetry present in the data.

APPENDIX

A Stationarity and persistence

Nelson (1990) discusses stationarity and persistence in the GARCH(1,1) model. I shall extend some of Nelson's results to the STARCH model. The general skedastic function (5) can be factorised as:

$$h_t = \omega + h_{t-1}(\alpha_1\,\epsilon_{t-1}^2 + \lambda\,\epsilon_{t-1}^2 F_{t-1} + \beta) \tag{47}$$

Repeatedly substituting for h_{t-i} in the above formula, we have for $t \geq 2$:

$$\begin{aligned} h_t \;=\; & h_0 \textstyle\prod_{i=1}^{t}[\alpha_1\,\epsilon_{t-i}^2 + \lambda\,\epsilon_{t-i}^2 F_{t-i} + \beta] \\ & + \omega[1 + \textstyle\sum_{k=1}^{t-1}\prod_{i=1}^{k}(\alpha_1\,\epsilon_{t-i}^2 + \lambda\,\epsilon_{t-i}^2 F_{t-i} + \beta)] \end{aligned} \tag{48}$$

This equation defines the conditional process of the h_t as starting from h_0. I shall suppose h_0 finite and strictly positive with probability one. Two questions will be addressed:

- When is the process of the h_t strictly stationary
- When does a shock given to the conditional process h_t decay as $t \to \infty$?

A first theorem, largely inspired from Nelson (1990), gives conditions for strict stationarity:

Theorem 3 *When $\omega = 0$, the conditional process of the h_t goes to zero almost surely if*

$$\mathrm{E}(\log[\alpha_1\,\epsilon_t^2 + \lambda\,\epsilon_t^2 F_t + \beta]) < 0 \tag{49}$$

and goes to infinity for the reverse case. When $\omega > 0$, the conditional process of the h_t is strictly stationary if (49) holds.

Proof: The proof follows directly from the appendix of Nelson (1990).

□

I shall now concentrate on a particular definition of persistence, persistence in L^1 which corresponds to Nelson's formula (36) and is related to the integrated GARCH of Engle and Bollerslev (1986). This type of persistence implies that the process is not weak stationary. I shall study persistence analysing the impact of an initial shock ϵ_0 on the conditional variance h_t.

Theorem 4 *A shock will be said to be persistent in L^1 unless*

$$lim_{t\to\infty}[\alpha_1 \mathrm{E}(\epsilon_t^2) + \lambda\,\mathrm{E}(\epsilon_t^2 F_t) + \beta]^{t-1} = 0 \tag{50}$$

Integration in the sense of Engle and Bollerslev corresponds to

$$\alpha_1 \mathrm{E}(\epsilon_t^2) + \lambda\,\mathrm{E}(\epsilon_t^2 F_t) + \beta = 1 \tag{51}$$

Proof: The impact of a normalised shock ϵ_0^2 on h_t is defined by

$$
\begin{aligned}
\frac{1}{\mathrm{E}(h_0)}\mathrm{E}\frac{\partial h_t}{\partial \epsilon_0^2} &= \mathrm{E}\left(\frac{\partial \prod_{i=1}^{t}[\alpha_1\,\epsilon_{t-i}^2 + \lambda\,\epsilon_{t-i}^2 F_{t-i} + \beta]}{\partial \epsilon_0^2}\right)\\
&= \mathrm{E}\left(\frac{\partial[\alpha_1\,\epsilon_0^2 + \lambda\,\epsilon_0^2 F_0 + \beta]}{\partial \epsilon_0^2}\right)\mathrm{E}\left(\prod_{i=1}^{t-1}[\alpha_1\,\epsilon_{t-i}^2 + \lambda\,\epsilon_{t-i}^2 F_{t-i} + \beta]\right)\\
&= (\alpha_1 + \lambda[\mathrm{E}(F_0) + \mathrm{E}(\epsilon_0^2\frac{\partial F_0}{\partial \epsilon_0^2})]) \times [\alpha_1 \mathrm{E}(\epsilon_t^2) + \lambda\,\mathrm{E}(\epsilon_t^2 F_t) + \beta]^{t-1}
\end{aligned}
\tag{52}
$$

using the factorisation (48). When $t \to \infty$

$$
[\alpha_1 \mathrm{E}(\epsilon_t^2) + \lambda\,\mathrm{E}(\epsilon_t^2 F_t) + \beta]^{t-1}
$$

goes to zero or to infinity depending if the inside bracket is lower or greater than one.

□

There is presumably no analytical result available to compute $\mathrm{E}(\epsilon_t^2 F_t)$, except when F is an indicator function. In this case, I can compute truncated moments for the two above cases: negative versus positive shocks and small versus big shocks. I shall assume that ϵ has a Normal distribution with zero mean and unit variance.

A.1 Negative and positive shocks

F is now zero when ϵ is negative and one otherwise. So as the distribution of ϵ is symmetric, negative and positive values are equally probable and:

$$
\mathrm{E}(\epsilon_t^2 F_t) = \int_{-\infty}^{0} \epsilon^2 f(\epsilon)\, d\epsilon = 1/2 \tag{53}
$$

The weak stationarity (non-persistence) condition becomes

$$
(\alpha_1 + \alpha_2)/2 + \beta < 1 \tag{54}
$$

and persistence is measured by α_1 for negative shocks and by α_2 for positive ones. For $\lambda = 0$, we recover the results of Engle and Bollerslev (1986).

A.2 Small and big shocks

F now is zero when $\epsilon \in [-c, c]$ and one otherwise. Because of the symmetry of the distribution of ϵ:

$$\mathrm{E}(\epsilon_t^2 F_t) = 1 - \int_{-c}^{c} \epsilon^2 f(\epsilon)\, d\epsilon$$

Assuming normality

$$\begin{aligned} \mathrm{E}(\epsilon_t^2 F_t) &= [1 - \mathrm{ERF}(c\frac{\sqrt{2}}{2}) + c\frac{\sqrt{2}}{\sqrt{\pi}} \exp(-\frac{c^2}{2})] \\ &= (1 - g(c)) \end{aligned} \tag{55}$$

where $\mathrm{ERF}(z)$ is the error function defined as the integral of the Gaussian function from zero to z. $g(c)$ is an increasing function of $c > 0$ taking values in $[0, 1]$. The weak stationarity condition becomes a not trivial function of c with:

$$(\alpha_1 + \lambda[1 - g(c)]) + \beta < 1 \tag{56}$$

For $c \to \infty$, we recover the usual condition for weak stationarity in linear GARCH with $\alpha_1 + \beta < 1$. For $c = 0$, we have the stationarity condition for the simple exponential transition function which is $\alpha_2 + \beta < 1$.

References

Bauwens L. and M. Lubrano, (1991), Bayesian Diagnostics for Heterogeneity, *Annales d'Economie et Statistique*, **20/21**, pp. 17-40.

Bauwens L. and M. Lubrano, (1998), Bayesian Inference in GARCH Models using the Gibbs Sampler, *The Econometrics Journal*, **1**, pp. C23-C46.

Bauwens L., M. Lubrano and J.F. Richard, (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford, Oxford University Press.

Bollerslev T, (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **31**(3), pp. 307-327.

Bollerslev T., R.Y. Chou and K.F. Kroner, (1992), ARCH modeling in finance, *Journal of Econometrics*, **52**(1-2), pp. 5-59.

Carlin B.P. and A.E. Gelfand, (1991), An iterative Monte Carlo method for non-conjugate Bayesian analysis, *Statistics and Computing*, **1**, pp. 119-128.

Engle R.F., (1982), Auto-regressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, **50**(4), pp. 987-1007.

Engle R.F. and T. Bollerslev, (1986), Modelling the Persistence of Conditional Variances, *Econometric Reviews*, **5**(1), pp. 1-50.

Engle R.F. and C. Mustafa, (1992), Implied ARCH Models for Option Prices, *Journal of Econometrics*, **52**(1-2), pp. 289-311.

Engle R.F. and V.K. Ng, (1993), Measuring and Testing the Impact of News on Volatility. *Journal of Finance*, **48**(5), pp. 1749-1778.

Engle R.F., A. Kane and J. Noh, (1996), Index Option Pricing with Stochastic Volatility and the Value of Accurate Variance Forecasts, *Review of Derivatives Research*, **1**(2), pp. 139-157.

Geweke J., (1993), Bayesian treatment of the independent Student-t linear model, *Journal of Applied Econometrics*, **8**(0) (Supplement), pp. 19-40.

Geweke J., (1994), Bayesian Comparison of Econometric Models, *Working Paper 532*, Research Department, Federal Reserve Bank of Minneapolis.

Glosten L.R., R. Jagannathan and D.E. Runkle, (1993), On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance* **48**(5), pp. 1779-1801.

Gonzales-Riviera G., (1996), Smooth Transition GARCH Models, Working Paper at Department of Economics, University of California at Riverside.

Granger C.W.J. and T. Terasvirta, (1993), *Modelling non-linear economic relationships*, Oxford, Oxford University Press.

Hagerud G.E. (1997) *A New Non-linear GARCH Model*, PhD Dissertation, Stockholm School of Economics.

Jansen E.S. and T. Terasvirta, (1996), Testing Parameter Constancy and Super Exogeneity in Econometric Equations, *Oxford Bulletin of Economics and Statistics*, **58**(4), pp. 735-763.

Kleibergen F. and H.K. van Dijk, (1993), Non-stationarity in GARCH Models: A Bayesian Analysis, *Journal of Applied Econometrics*, **8**(0), Supplement, pp. S41- S61.

Lubrano M., (2000), Bayesian Analysis of Non-linear Time Series with Models with a Threshold, in W.A. Barnett, D.F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A. Würtz (eds), *Non-linear Econometric Modelling*, Cambridge University Press, pp. 79-118.

Luukkonen R., P. Saikkonen and T. Terasvirta, (1988), Testing Linearity against Smooth Transition Autoregression, *Biometrika*, **75**, pp. 491-499.

Nelson D.B., (1990), Stationarity and persistence in the GARCH(1,1) model, *Econometric Theory*, **6**(3), pp. 318-334.

Nelson D.B., (1991), Conditional heteroskedasticity in asset returns: A new approach, *Econometrica*, **59**(2), pp. 347-370.

Noh J., R.F. Engle and A. Kane, (1995), Forecasting Volatility and Option Prices of the SP500 Index, In Engle Robert F., ed. *ARCH: Selected readings, Advanced Texts in Econometrics*, Oxford and New York: Oxford University Press, pp. 314-331.

Osiewalski J. and A. Welfe, (1998), The Wage-Price Mechanism: An Endogenous Switching Model, *European Economic Review*, **42**(2), pp. 365-374.

Pagan A., (1996), The econometrics of financial markets, *Journal of Empirical Finance*, **3**, pp. 15-102.

Pagan A. and G.W. Schwert, (1990), Alternative Models for Conditional Stock Volatility, *Journal of Econometrics*, **45**(1-2), pp. 267-290.

Robert C.P. and K.L. Mengersen ,(1995), Reparameterisation Issues in Mixture Modelling and their Bearing on the Gibbs Sampler, CREST-INSEE Discussion Paper No 9538.

Susmel R. and R.F. Engle , (1994), Hourly Volatility Spillovers between International Equity Markets, *Journal of International Money and Finance*, **13**(1), pp. 3-25.

Terasvirta T., (1994) Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models, *Journal of the American Statistical Association*, **89**(425), pp. 208-218.

Zakoian J.M., (1994), Threshold Heteroskedastic Models, *Journal of Economic Dynamics and Control*, **18**(5), pp. 931-955.

Zellner A., (1975), Bayesian Analysis of Regression Error Terms, *Journal of the American Statistical Association*, **70**, pp. 138-144.