Non-Stationarity in GARCH Models: A Bayesian Analysis
Author(s): F. Kleibergen and H. K. Van Dijk
Reviewed work(s):
Source: *Journal of Applied Econometrics,* Vol. 8, Supplement: Special Issue on Econometric
Inference Using Simulation Techniques (Dec., 1993), pp. S41–S61
Published by: John Wiley & Sons
Stable URL: http://www.jstor.org/stable/2285074
Accessed: 20/07/2012 11:03

NON-STATIONARITY IN GARCH MODELS: A BAYESIAN ANALYSIS

F. KLEIBERGEN AND H. K. VAN DIJK

Econometric Institute and Tinbergen Institute, Oostmaaslaan 950-952, 3063 DM Rotterdam, The Netherlands

SUMMARY

First, the non-stationarity properties of the conditional variances in the GARCH(1, 1) model are analysed using the concept of infinite persistence of shocks. Given a time sequence of probabilities for increasing/decreasing conditional variances, a theoretical formula for quasi-strict non-stationarity is defined. The resulting conditions for the GARCH(1, 1) model are shown to differ from the weak stationarity conditions mainly used in the literature. Bayesian statistical analysis using Monte Carlo integration is applied to analyse both stationarity concepts for the conditional variances of the US 3-month treasury bill rate. Interest rates are known for their weakly non-stationary conditional variances but, using a quasi-strict stationarity measure, it is shown that the conditional variances are likely to be stationary. Second, the level of the treasury bill rate is analysed for non-stationarity using Bayesian unit root methods. The disturbances of the GARCH model for the treasury bill rate are *t*-distributed. It is shown that the unit root parameter is negatively correlated with the degrees-of-freedom parameter. Imposing normally distributed disturbances leads therefore to underestimation of the non-stationarity in the level of the treasury bill rate.

1. INTRODUCTION

The literature on so-called Generalized AutoRegressive Conditional Heteroscedasticity (GARCH) models has grown rapidly since the introduction of ARCH models by Engle (1982) (see, for example, the surveys by Bollerslev *et al.*, 1992; Bera and Higgins, 1992). However, some features of GARCH models are not well known, in particular, possible non-stationarity of the conditional variance in the GARCH(1, 1) model. The distinction between weak and strict stationarity plays a crucial role in this respect. We note that the odd kind of implications of the weak stationarity concept for the conditional variances have been commented upon by Geweke (1986), while Nelson derived (1990) derived the strict stationarity conditions.

In this paper we analyse the difference between strict and weak stationarity using the concept of a sequence of probabilities of increasing or decreasing conditional variances. This helps to give an intuitive explanation why the two conditions differ in case of the GARCH(1, 1) model. If one defines the key feature of non-stationarity (explosiveness) by the infinite persistence of shocks, one expects a sequence of probabilities of increasing conditional variances to exhibit an upward-sloping behaviour over the number of periods (and to converge to 1 in the limit). The resulting quasi-strict stationarity condition (definition 1) is shown to give different parameter restrictions for the GARCH(1, 1) model from the weak stationarity condition. It is interesting to note that the weak stationarity conditions are far more restrictive than the quasi-strict stationarity conditions for the GARCH(1, 1) model. The two different stationarity conditions are discussed in Section 2.

Due to the complicated non-linearity appearing in the quasi-strict stationarity condition, a classical statistical analysis of this condition will be hard to pursue. We therefore rely on Bayesian methods which approximate the posterior of the quasi-strict stationarity condition using simulation methods. As an example in Section 3, we analyse an interest rate series, the US 3-month treasury bill rate. Some interest rate series are known to have conditional variances which are weakly non-stationary (see Bollerslev *et al.*, 1992). The quasi-strict stationarity conditions have not, however, been analysed for these series. Given the questionable appropriateness of the weak stationarity concept for the GARCH(1, 1) model, we prefer the stationarity conclusions to be based on the quasi-strict stationarity conditions.

Non-stationarity in the conditional variances is not the only possible source of non-stationarity for interest rate series. The level of the treasury bill rate may also be non-stationary. In Section 4 we analyse non-stationarity in the level of the series by conducting a Bayesian unit root analysis. The volatile behaviour of the disturbances contradicts the assumption of normally distributed disturbances. The disturbances are therefore modelled as independent Student *t*-distributed random variables with unknown common degrees of freedom. By modelling the disturbances using an independent Student distribution, an interesting phenomenon appeared, i.e. the negative correlation of the degrees of freedom with the unit root parameter. Imposition of normality underestimates the non-stationarity of the interest rate level as a consequence. Robustness arguments can be used to explain this phenomenon. Finally, in the fifth section some conclusions are drawn.

2. PERSISTENCE AND STATIONARITY IN GARCH MODELS

In Bollerslev (1986) it is shown that the GARCH model belongs to the class of AutoRegressive Moving Average (ARMA) models. Consider a stochastic process $\{x_t\}_{t=1}^{\infty}$ defined by a certain statistical model with disturbances $\{\varepsilon_t\}_{t=1}^{\infty}$ with conditional variances, $h_t = E_t(\varepsilon_t^2)$, which are generated by a GARCH(1, 1) model. The relationship between this GARCH(1, 1) model and an ARMA (1, 1) model reads

$$h_t = \gamma + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} \Leftrightarrow \tag{1}$$

$$\varepsilon_t^2 = \gamma + (\alpha + \beta)\varepsilon_{t-1}^2 - \beta\eta_{t-1} + \eta_t \tag{2}$$

where α and $\beta \geqslant 0$, $\gamma > 0$, $\eta_t = \varepsilon_t^2 - h_t = h_t(z^2 - 1)$, and z is a standardized random variable with mean 0 and scaling factor 1; for example, $z \sim n(0, 1)$. The difference between the model in equation (2) and a standard ARMA model refers to the parameters α and β and the disturbances η_t. The parameters α and β are restricted to be non-negative, which has some consequences for the stationarity conditions, and although the disturbances η_t have mean 0, they are clearly not white noise because of their time-varying asymmetric probability density functions. Bollerslev (1986) and Engle and Bollerslev (1986) discuss the weak stationarity conditions of the restricted ARMA (GARCH) model under the non-negativity of the parameters. One can show, however, that, due to the asymmetric probability density function of the disturbances, the strict stationarity condition differs from this weak stationarity condition.

To show this difference a quasi-strict stationarity definition will be used. Strict stationary stochastic processes are defined by the invariance of the joint distribution of several successive realizations of the stochastic process with respect to the time period of the initial realization (see Box and Jenkins, 1976). Note that strict stationarity refers to the joint probability density function while weak stationarity is defined in terms of the moments of the process. If the

disturbances are normally distributed, the two concepts coincide. The functional form of the joint probability density function of a set of successive realizations from a stochastic process is often very hard to analyse. We therefore use a convenient definition of quasi-strict stationarity (non-explosiveness) based on probability measures.

Definition 1. A stochastic process $\{x_t\}_{t=1}^{\infty}$ is quasi-strict stationary if, for every t,

$$\Pr[|x_{t+j}| > |x_t|] = 1 - \varepsilon_j, \quad \varepsilon_j > \varepsilon > 0, \ j = 1, \dots, \infty$$

The quasi-strict (tail probability) definition guarantees that x_t is measurable, i.e. it prevents $|x_t|$ from converging to infinity with probability 1, so the series are not explosive, and it prevents the probabilities from being time dependent. The definition of strict stationarity used in the literature is slightly different from the tail probability definition stated above, because it considers all possible values of the stochastic process $\{x_t\}_{t=1}^{\infty}$, while the definition stated above considers only values of x_t for which $|x_{t+j}| > |x_t|$, $j = 1, \dots, \infty$. The tail probability condition guarantees that explosive processes are quasi-strict non-stationary. It also implies that stochastic processes with time-varying parameter models are quasi-strict non-stationary because the assumed independence between the ε_j terms and the time period, t, is violated. As a consequence, for many processes the conditions for strict and quasi-strict stationarity will coincide.

Before we discuss the difference between the quasi-strict stationarity and weak stationarity concepts for GARCH models we show that both definitions have the same implications for ARMA models. To be able to analyse both stationarity concepts we need a definition of weak stationarity (see Box and Jenkins, 1976).

Definition 2. A stochastic process $\{x_t\}_{t=1}^{\infty}$ is weakly stationary if, for every t and j, $j = 1, \dots, \infty$,

$$-\infty < \mathrm{E}(x_t) = \mu < \infty, \quad -\infty < \mathrm{cov}(x_t, x_{t+j}) = \gamma_j < \infty$$

Application of the weak stationarity definition to an ARMA(p, q) model with i.i.d. disturbances is quite straightforward. Consider

$$\rho(L)x_t = \alpha(L)\varepsilon_t \tag{3}$$

where $\rho(L)$, $\alpha(L)$ are finite-order lag polynomials specified by

$$\rho(z) = 1 - \sum_{i=1}^{p} \rho_i z^i \qquad \alpha(z) = 1 + \sum_{i=1}^{q} \alpha_i z^i \tag{4}$$

and $\mathrm{var}(\varepsilon_t) = \sigma^2$. The autocovariances can be calculated by assuming that $\rho(L)$ is invertible, which allows the construction of the implied infinite-order Moving Average (MA) process:

$$x_t = c(L)\varepsilon_t = \rho(L)^{-1}\alpha(L)\varepsilon_t = \sum_{i=0}^{j-1} c_i\varepsilon_{t-i} + C_j x_{t-j} \tag{5}$$

where

$$c(z) = \sum_{i=0}^{\infty} c_i z^i \quad \text{and} \quad C_j = \sum_{i=0}^{\infty} c_i c_{i+j} \bigg/ \sum_{i=0}^{\infty} c_i^2, \qquad j = 0, \dots, \infty$$

The autocovariances then become

$$\gamma_j = \mathrm{cov}(x_{t+j}, x_t) = \sigma^2 \sum_{i=0}^{\infty} c_i c_{i+j} = \sigma^2 C_j \left(\sum_{i=0}^{\infty} c_i^2\right) \qquad j = 0, \dots, \infty \tag{6}$$

The autocovariances are independent from the observation time period t, implying that univariate ARMA processes are weakly stationary when the parameters γ_j, $j = 0, ..., \infty$ are finite. The parameters $\gamma_j = C_j(\Sigma_{i=0}^{\infty} c_i^2)$ are finite when both the polynomial $\alpha(z)$ has finite parameters and the roots of the polynomial $\rho(z)$ lie outside the unit circle. If the roots of the polynomial $\rho(z)$ lie within the unit circle, the parameters c_i converge to infinity when i increases, which results in an infinite value of γ_j.

To analyse the quasi-strict stationarity measure, we substitute equation (5) in definition 1. This yields

$$\Pr[|x_{t+j}| > |x_t|] = \Pr\left[\left|\sum_{i=0}^{j-1} c_i \varepsilon_{t+j-i} + C_j x_t\right| > |x_t|\right] \tag{7}$$

Taking the limit for $j \to \infty$ gives

$$\lim_{j \to \infty} \Pr\left[\left|\sum_{i=0}^{j-1} c_i \varepsilon_{t+j-i} + C_j x_t\right| > |x_t|\right] = 1 \Leftrightarrow \lim_{j \to \infty} C_j = \infty \tag{8}$$

The disturbances ε_t have a symmetric probability density function with mode 0. Therefore the probability density function of $\Sigma_{i=0}^{j-1} c_i \varepsilon_{t+j-i}$ is also symmetric with mode 0. The limiting behaviour of the probability in equation (7) depends solely on the convergence behaviour of C_j when $j \to \infty$ as a consequence. When C_j converges to infinity, the probability measure will converge to 1 and the stochastic process is non-stationary. C_j converges to infinity when the roots of polynomial $\rho(z)$ lie within or on the unit circle and/or when the polynomial $\alpha(z)$ has infinite parameters. Equations (7) and (8) also show that the probability measure is independent from the time period t. Therefore the only condition which needs to be verified is the boundedness of the probabilities. The quasi-strict stationarity condition correspond to the condition derived for the weak stationarity definition. As a consequence, quasi-strict stationarity corresponds with weak stationarity in ARMA models.

The weak stationarity condition of the GARCH$(1, 1)$ model is constructed using a multiplicative specification of the GARCH$(1, 1)$ model, which is also convenient for the construction of the quasi-strict stationarity condition (see also Bollerslev, 1986; Nelson, 1990):

$$h_t = \gamma + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \tag{9}$$

$$= h_0 \prod_{i=1}^{t} (\beta + \alpha z_{t-i}^2) + \gamma\left(1 + \sum_{k=1}^{t-1} \prod_{i=1}^{k} (\beta + \alpha z_{t-i}^2)\right) \tag{10}$$

$$= h_0 G_t + \gamma \sum_{k=0}^{t-1} G_k \tag{11}$$

where $z_t = \varepsilon_t / h_t^{1/2}$, $G_j = G_{j-1}(\beta + \alpha z_{t-j}^2)$, $G_0 = 1$. The mean of the conditional variances in a GARCH$(1, 1)$ model can directly be calculated from equation (11). Assuming normality, one has

$$E(G_k) = E(G_{k-1}(\beta + \alpha z_{t-k}^2)) = E(G_{k-1})E(\beta + \alpha z_{t-k}^2) = (\alpha + \beta)^k \tag{12}$$

so that the unconditional variance becomes

$$E(h_t) = \gamma \sum_{k=0}^{\infty} E(G_k) = \gamma \sum_{k=0}^{\infty} (\alpha + \beta)^k = \gamma/(1 - \alpha - \beta) \qquad \alpha + \beta < 1$$

$$= \infty \qquad \alpha + \beta \geqslant 1 \tag{13}$$

To calculate the variance of the conditional variances of the GARCH$(1, 1)$ model we need to know the third- and fourth-order moments of the disturbances. Once these moments are

known it is again straightforward to calculate these variances and covariances. Assuming normality, one can derive that

$$E(G_k^2) = E(G_{k-1}^2)E(\beta + \alpha z_{k-1}^2)^2 = (\beta^2 + 2\alpha\beta + 3\alpha^2)^k \tag{14}$$

$$E(G_k G_j) = E\left(G_j^2 \prod_{i=1}^{k-j} (\beta + \alpha z_{j+i}^2)\right) = (\beta + \alpha)^{k-j}(\beta^2 + 2\alpha\beta + 3\alpha^2)^j \tag{15}$$

$$E(h_t^2) = \gamma^2 E\left(\sum_{k=0}^{\infty} G_k\right)^2 = \gamma^2 \left[\sum_{k=0}^{\infty} E(G_k^2) + 2 \sum_{k=0}^{\infty} \sum_{j=0}^{k-1} E(G_k G_j)\right]$$
$$= \gamma^2(1 + \alpha + \beta)/((1 - (\beta^2 + 2\alpha\beta + 3\alpha^2))(1 - (\alpha + \beta)))$$
$$(\beta^2 + 2\alpha\beta + 3\alpha^2) < 1$$
$$= \infty \quad (\beta^2 + 2\alpha\beta + 3\alpha^2) \geqslant 1 \tag{16}$$

Therefore the variance of the conditional variances is given by

$$\text{var}(h_t) = E(h_t^2) - (E(h_t)^2)$$
$$= 2\gamma^2\alpha^2/((1 - (\beta^2 + 2\alpha\beta + 3\alpha^2))(1 - (\alpha + \beta))^2) \quad (\beta^2 + 2\alpha\beta + 3\alpha^2) < 1$$
$$= \infty \quad (\beta^2 + 2\alpha\beta + 3\alpha^2) \geqslant 1 \tag{17}$$

In the case of more general error distributions such as the Student t, the mean and variance of h_t can be calculated by using the moments of a standardized t-distributed random variable, z:

$$E(h_t) = \gamma/(1 - \alpha E(z^2) - \beta) \quad \alpha E(z^2) + \beta < 1 \tag{18}$$

$$\text{var}(h_t) = \gamma^2 [E(z^4) - E(z^2)]\alpha^2/[(1 - (\beta^2 + 2E(z^2)\alpha\beta + E(z^4)\alpha^2))(1 - (E(z^2)\alpha + \beta))^2]$$
$$\beta^2 + 2E(z^2)\alpha\beta + E(z^4)\alpha^2 < 1 \tag{19}$$

The covariance of h_t and h_j can be calculated using equations (11) and (15):

$$\text{cov}(h_t, h_j) = \text{cov}\left(h_0 G_t + \gamma \sum_{k=0}^{t-1} G_k, h_0 G_j + \gamma \sum_{k=0}^{j-1} G_k\right)$$
$$= (\alpha E(z^2) + \beta)^{t-j} \text{var}(h_t) \quad t > j \tag{20}$$

If $(1 - (\beta^2 + 2E(z^2)\alpha\beta + E(z^4)\alpha^2)) > 0$, the variances and covariances are finite and time independent. The GARCH(1, 1) conditional variance process is weakly stationary in that case. The stationarity properties indicated by the weak stationarity condition can be quite restrictive in practice in the sense that even for values of α and β which indicate weak non-stationarity, the process behaves quite regularly. The probability of a decreasing variance in the next period(s) exceeds the probability of an increasing variance even for values of α and β, which are said to be weakly non-stationary. If the probability of an increasing variance is exceeded by that of a decreasing variance, shocks to the variance are not likely to persist for long. If this property holds for non-stationary series, one has a strange kind of definition of non-stationarity because one of the key conditions of non-stationarity, infinite persistence of shocks, is violated. The probability of a decreasing conditional variance can be analysed using

$$\Pr[h_{t+1} < h_t] = \Pr[\gamma + \alpha\varepsilon_t^2 + \beta h_t < h_t]$$
$$= \Pr[\varepsilon_t^2/h_t < ((1 - \beta) - \gamma/h_t)/\alpha] \tag{21}$$
$$= \Pr[\varepsilon_t^2/h_t < 1 - \gamma/\alpha h_t] \quad \alpha + \beta = 1 \tag{22}$$
$$= 0 \cdot 68 \quad \gamma = 0, \text{ normal dist.} \tag{23}$$

The probability measure in equation (21) also allows one to construct an upper bound and the unconditional expectation of this probability measure:

$$\mathcal{E}_0(\Pr[h_{t+1} < h_t]) = \mathbf{E}_0(\Pr[\varepsilon_t^2/h_t < ((1 - \beta) - \gamma/h_t)/\alpha])$$

$$= \Pr[\varepsilon_t^2/h_t < 1] \qquad \alpha \mathbf{E}(z^2) + \beta < 1$$

$$= \Pr[\varepsilon_t^2/h_t < (1 - \beta)/\alpha] \qquad \alpha \mathbf{E}(z^2) + \beta \geqslant 1 \qquad (25)$$

$$= 0 \cdot 68 \qquad \alpha + \beta < 1, \text{ normal} \qquad (26)$$

$$\Pr[h_{t+1} < h_t] = \Pr[\varepsilon_t^2/h_t < ((1 - \beta) - \gamma/h_t)/\alpha] \qquad (27)$$

$$\leqslant \Pr[\varepsilon_t^2/h_t < (1 - \beta)/\alpha] \qquad (28)$$

The upper bound on the probability of a decreasing variance will be tight for large values of the conditional variance h_t and/or small values of the parameter γ. Both these phenomena are typical for series with weakly non-stationary conditional variances, as will be shown later. This is also in agreement with the unconditional expected value of the probability of a decreasing variance which equals the upper bound in case of weakly non-stationary conditional variances.

As an example of a series with possible non-stationary conditional variances we use the 3-month US treasury bill rate from January 1957 to April 1989 (388 observations). Interest rates are known for their weakly non-stationary conditional variances (see Bollerslev *et al.*, 1992). The analysed interest rate series is shown in Figure 1. The assumption of GARCH(1, 1) conditional variances is often not sufficient to explain the heteroscedastic and leptokurtic nature of the disturbances of exchange rates and interest rates (see Engle and Bollerslev, 1986; Spanos, 1991). Following these authors, we obtain Maximum Likelihood (ML) estimators of the parameters of a model incorporating independent *t*-distributed disturbances with unknown common degrees of freedom by means of numerical optimization. Using the Akaike information criterium an AutoRegressive model of order 1 (AR(1)) with GARCH(1, 1) conditional variances and *t*-distributed disturbances is selected to describe the behaviour of the
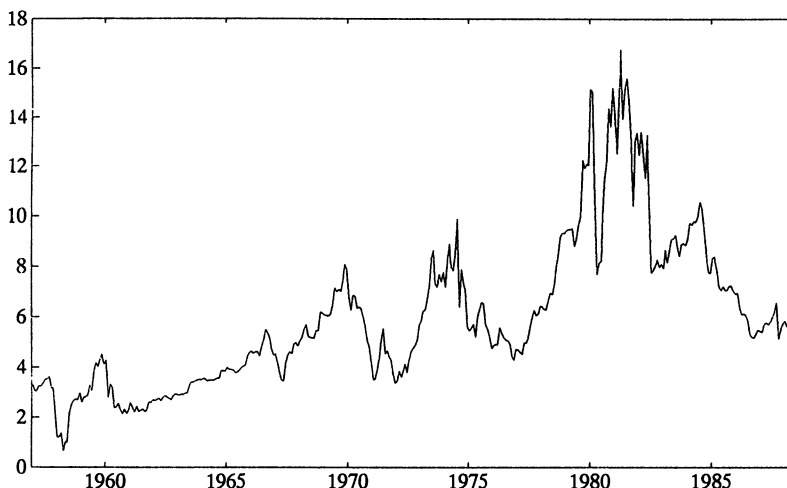

Figure 1. US 3-month treasury bill rate from January 1957 to April 1989

series $(h_0 = \Sigma_{t=1}^{T} \varepsilon_t^2 / T)$:

$$y_t = 0\cdot016 + 1\cdot00 y_{t-1} + \varepsilon_t \qquad\qquad \varepsilon_t = 1, \ldots, 387$$
$$(1\cdot22) \quad (398)$$

$$h_t = 0\cdot00038 + 0\cdot013\varepsilon_{t-1}^2 + 0\cdot77 h_{t-1} \qquad\qquad \varepsilon_t \sim t(0, h_t, \lambda), \quad \lambda = 3\cdot00$$
$$(0\cdot82) \quad\quad (3\cdot3) \quad\quad\quad (16\cdot8) \qquad\qquad\qquad\qquad\qquad (5\cdot62)$$

$$p(\varepsilon_t) = \frac{\Gamma(\frac{1}{2}(\lambda + 1))}{\Gamma(\frac{1}{2}\lambda)\Gamma(\frac{1}{2})} (\lambda h_t)^{-1/2} \left| 1 + \frac{\varepsilon_t^2}{\lambda h_t} \right|^{-1/2(\lambda+1)}$$

$$\text{LR}_{\text{nor-dis}} = 20\cdot3 \qquad\qquad\qquad\qquad \text{LR}_{\text{cons-var}} = 49\cdot8 \qquad\qquad (29)$$

The Likelihood Ratio (LR) statistics for testing the hypotheses of a large number of degrees of freedom which approximately corresponds with normally distributed disturbances ($20\cdot3$) and constant variances ($49\cdot8$) show that models which neglect the heteroscedastic or Student t properties of the disturbances are strongly rejected. The parameter estimates indicate that the hypothesis of weakly non-stationary conditional variances cannot be rejected because $1 - \alpha E(z^2) - \beta = 1 - \alpha\lambda/(\lambda - 2) - \beta = -0\cdot16$. The hypothesis of non-stationarity in the level of the series is not rejectable either because of the estimated value of AR(1) parameter ρ of $1\cdot0$. These properties make this treasury bill series suitable for a stationarity analysis of the conditional variance (Section 3) and of the level of the series (Section 4).

Given the estimated values of the parameters of the interest rate model it is possible to calculate the conditional variances and probabilities of a decrease in the conditional variance. Figure 2 contains these conditional variances together with the probabilities of a decrease of the conditional variance in the next period and the upper bound on this probability. Figure 3 shows the empirical probabilities of an increase in the conditional variance in the next j periods for $j = 1, \ldots, 100$. The high values of the probability of a decrease of the conditional variance in the next period shown in Figure 2 are not a sound indication of stationarity because of the increase in the empirical probability of an increasing conditional variance after one period (Figure 3). The empirical probabilities in Figure 3 do not give a clear indication of stationarity
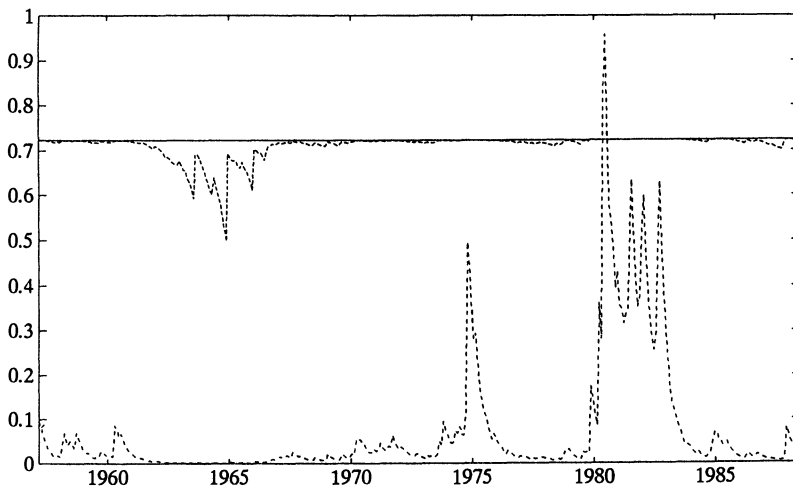

Figure 2. Probability decreasing variance next period (—), upper bound (– –) and ($0\cdot25\times$) conditional variances ($\cdots$)
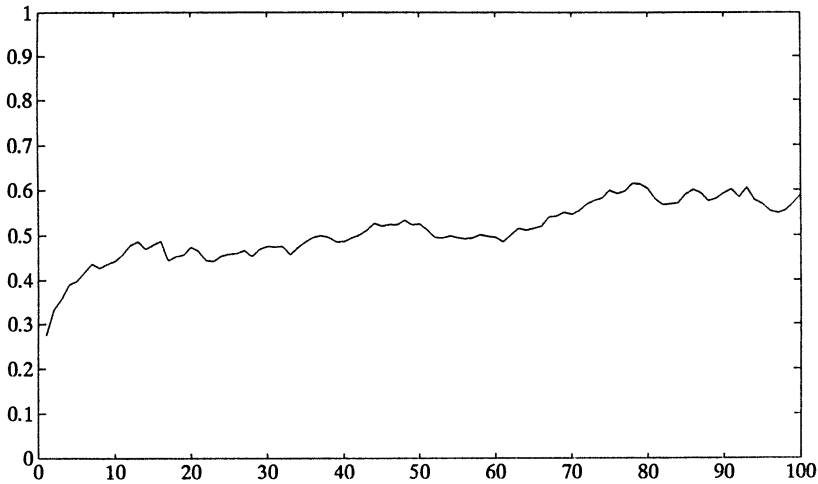
Figure 3. Probability of an increasing variance in the next j periods

or non-stationarity either, because the probabilities keep lingering around $0 \cdot 5$ and do not show a clearcut upward- or downward-sloping behaviour. The value of $1 - \alpha E(z^2) - \beta(-0 \cdot 16)$ clearly indicates weak non-stationarity, however, while Figures 2 and 3 are quite indecisive between stationarity or non-stationarity of the conditional variances. The weak stationarity condition seems to overestimate the true amount of non-stationarity in the analysed interest rate conditional variances as a consequence.

Figures 2 and 3 show that analysing non-stationarity by using the weak stationarity condition overestimates the amount of non-stationarity in the conditional variances. This results from the fact that the G_k process does not converge to its mean, $(\alpha E(z^2) + \beta)^k$, as implicitly assumed in the weak stationarity condition, but to a value which is well below its mean for many values of α and β. It can be shown that the G_k process converges to 0 for a region of values of α and β which exceeds the region, $\alpha E(z^2) + \beta < 1$, which was used in the derivation of the weak stationarity concept.

Theorem 1. The stochastic process $\{G_k\}_{k=0}^{\infty}$, $G_0 = 1$, $G_k = G_{k-1}(\beta + \alpha z_k^2)$, converges to 0 when $k \to \infty$ when $\varepsilon(\ln(\beta + \alpha z^2)) < 0$, elsewhere the process converges to ∞ (see also Nelson, 1990).

Proof:

$$\lim_{k \to \infty} G_k = \lim_{k \to \infty} \prod_{j=1}^{k} (\beta + \alpha z_j^2)$$

$$= \lim_{k \to \infty} \exp\left[\sum_{j=1}^{k} \ln(\beta + \alpha z_j^2)\right]$$

$$= \exp\left[\lim_{k \to \infty} \sum_{j=1}^{k} \ln(\beta + \alpha z_j^2)\right]$$

(Using Slutsky's theorem: $g(x)$ continuous, $\text{plim}_{T \to \infty} g(x_T) = g(\text{plim}_{T \to \infty} x_T)$)

$$= \exp\left[\lim_{k \to \infty} k E(\ln(\beta + \alpha z^2))\right]$$

$$\left(\text{Strong law of large numbers:} \ \plim_{k \to \infty} k^{-1} \sum_{j=1}^{k} z_j = \varepsilon(z) \right)$$

$$= 0 \qquad \mathrm{E}(\ln(\beta + \alpha z^2)) < 0$$

$$= \infty \qquad \mathrm{E}(\ln(\beta + \alpha z^2)) \geqslant 0$$

Using a Taylor approximation of $\ln(\beta + \alpha z^2)$ one can easily show that the region where $\mathrm{E}(\ln(\beta + \alpha z^2)) < 0$ includes one where $1 - \beta - \alpha \mathrm{E}(z^2) > 0$, i.e. the region with 'finite unconditional variances'. This can be derived using a first-order Taylor approximation of $\mathrm{E}(\ln(\beta + \alpha z^2)) = \mathrm{E}[(\beta + \alpha z^2 - 1) - \frac{1}{2}(\xi)^{-2}(\beta + \alpha z^2 - 1)^2] \leqslant \mathrm{E}((\beta + \alpha z^2 - 1)$, where ξ is a scalar between 0 and $\beta + \alpha z^2 - 1$ (Jensen inequality). So $\mathrm{E}(\ln(\beta + \alpha z^2)) \leqslant \mathrm{E}(\beta + \alpha z^2 - 1)$ and the region where $1 - \beta - \alpha \mathrm{E}(z^2) > 0$ is included in the region where $\mathrm{E}(\ln(\beta + \alpha z^2)) < 0$. Figure 4 shows the region of values of α and β where $\mathrm{E}(\ln(\beta + \alpha z^2)) < 0$ for different distributions of the disturbances. These regions of values of (α, β) for which G_k converges to 0 considerably exceed the region of 'finite unconditional variances', especially when one bears in mind that in case of Cauchy disturbances the unconditional variance is always 'infinite'. The existence of values of (α, β) for which $\mathrm{E}(\ln(\beta + \alpha z^2)) < 0$ but $1 - \beta - \alpha \mathrm{E}(z^2) < 0$ shows the inappropriateness of the weak stationarity concept to the GARCH(1, 1) model. In the derivation of the weak stationarity condition the convergence of the G_k terms towards 0 was determined through the $(\beta + \alpha z_k^2)$ variables independently. The convergence of the G_k terms is accelerated through the multiplicative effect of the asymmetry of the probability density function of $(\beta + \alpha z_k^2)$ on the G_k terms. The weak stationarity condition neglects this multiplicative effect which the different random variables $(\beta + \alpha z_k^2)$ have on one another. The stationary region resulting from the weak stationarity concept almost equals but is still exceeded by the region where $\Pr[(\beta + \alpha z_k^2) < 1] > 0 \cdot 5$ ($\Leftrightarrow 0 \cdot 83\alpha + \beta < 1$ in case of normality) namely, the region where the probability of a decrease in the conditional variance in the next period exceeds $0 \cdot 5$ or, put differently, the region where $\Pr[G_k > G_{k+1}] > 0 \cdot 5$. As a consequence, the weak stationarity condition neglects the multiplicative effect. This effect
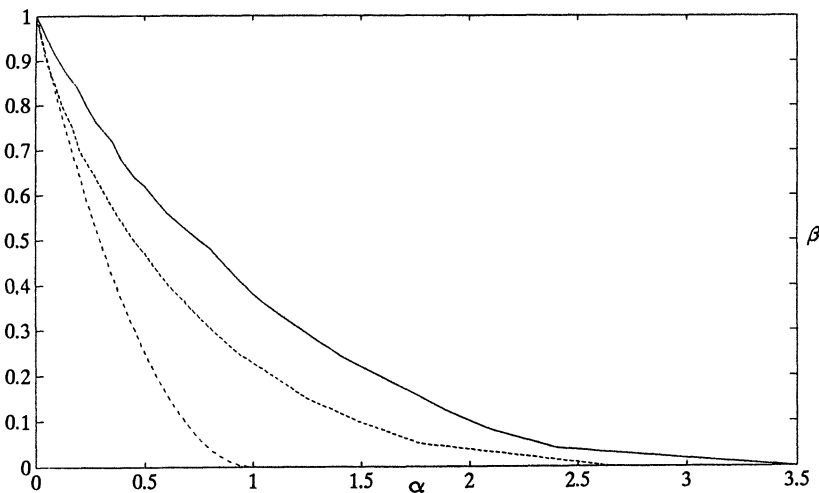

Figure 4. Region where $\mathrm{E}(\ln(\beta + \alpha z^2)) < 0$ or stationary conditional variances for $z \sim n(0, 1)$ (—), $z \sim$ Cauchy$(0, 1)$ (···) and $z \sim t(0, 1, 3)$ (— —)

combined with the asymmetry of the probability density function of z^2 accelerates the convergence of the G_k terms towards 0 compared with the independent case.

Essential to the concept of stationary conditional variances is that the G_k terms should converge to 0 at such a speed that their sum is finite. The negativity of $E(\ln(\beta + \alpha z^2))$ suffices to secure the finiteness of the sum of the G_k terms and to ensure quasi-strict stationarity as a consequence.

Theorem 2. The conditional variances of a GARCH(1, 1) process are stationary such that

$$\forall t \; \Pr[h_{t+j} > h_t] = 1 - \varepsilon_j, \; \varepsilon_j > 0, \qquad j = 1, \ldots, \infty$$

$$\text{if and only if } E(\ln(\beta + \alpha z^2)) < 0$$

Proof:

$$\Pr[h_{t+j} > h_t] = \Pr\left[h_t G_j + \gamma \sum_{k=0}^{j-1} G_k > h_t\right] \qquad (h_t \geqslant \gamma)$$

$$\leqslant \Pr\left[h_t \sum_{k=1}^{j} G_k > h_t - \gamma\right]$$

$$= \Pr\left[\sum_{k=1}^{j} G_k > 1 - \gamma/h_t\right]$$

$$= \Pr\left[\sum_{k=1}^{j} \exp\left(\sum_{i=1}^{k} \ln(\beta + \alpha z_{t+j-i}^2)\right) > 1 - \gamma/h_t\right]$$

$$\lim_{j \to \infty} \Pr[h_{t+j} > h_t] = 1 \qquad E(\ln(\beta + \alpha z^2)) \geqslant 0 \quad \text{(using Slutsky's theorem,}$$
$$\text{see theorem 1)}$$

$$< 1 - \varepsilon \qquad E(\ln(\beta + \alpha z^2)) < 0$$

In Nelson (1990) the strict stationarity conditions for the GARCH(1, 1) model are derived and they coincide with the quasi-strict conditions in theorem 2. In the next section the stationarity measure $E(\ln(\beta + \alpha z^2))$ will be applied to analyse the amount of non-stationarity present in the conditional variances of the analysed interest rate.

The differences between the weak and quasi-strict stationarity conditions in the GARCH(1, 1) model show that one has to be careful with the construction of non-stationarity conditions in non-standard models, i.e. models which are not in a linear form or those containing disturbances with asymmetric probability density functions. In constructing these non-stationarity conditions it is also better to use quasi-strict stationarity measures such as definition 1, because the equality of the weak and strict stationarity conditions in ARMA models happens to be the result of the symmetry of the probability density function of the disturbances and the linearity of the ARMA model and does not generalize directly to other models.

3. NON-STATIONARITY ANALYSIS OF THE CONDITIONAL VARIANCE OF THE US TREASURY BILL RATE

The more formal statistical analysis of the (non) stationarity of the conditional variances of the AR(1)-GARCH(1, 1) model for the US treasury bill series will be conducted using Bayesian

statistical analysis. Performing a Bayesian analysis means that one has to construct and calculate the posteriors of the parameters. Using a flat prior, the marginal posteriors of the parameters of an AR(1)-GARCH(1, 1) model with t-distributed disturbances are analytically intractable. These posteriors have to be calculated using numerical techniques therefore. A popular method of numerical calculation of the posteriors is the so-called Gibbs sampling technique (see Gelfand and Smith, 1990). For this technique to be both elegant and efficient one has to know the analytical form of the conditional posteriors of each set of parameters given the remaining parameters. The presence of the GARCH conditional variances obscures the analytical construction of these conditional posteriors. If one does not have these GARCH conditional variances the Gibbs sampling technique is elegant to apply even in the presence of t-distributed disturbances (see Geweke, 1992). The Gibbs sampling technique is still applicable with GARCH conditional variances but one has to use approximating functions of certain conditional posteriors. Instead of approximating certain different conditional posteriors we choose to approximate the joint posterior of all parameters by a multivariate t density with one degree of freedom and to perform Importance Sampling (see Kloek and van Dijk, 1978; van Dijk and Kloek, 1980; Geweke, 1989b). Importance Sampling using models with ARCH conditional variances and normally distributed disturbances has already been applied by Geweke (1989a). The program SISAM (see Hop and van Dijk, 1992) has been used to calculate the marginal posteriors of the different parameters of interest.

Using the SISAM program and flat uniform priors for all parameters, the following means and 't-values' of the marginal posteriors of the model in equation (29) were calculated:

$$
\begin{aligned}
&y_t = 0 \cdot 022 + 1 \cdot 00 y_{t-1} + \varepsilon_t && t = 1, \ldots, 387 \\
&\quad (0 \cdot 68) \quad\ (129) \\
&h_t = 0 \cdot 0014 + 0 \cdot 18\varepsilon_{t-1}^2 + 0 \cdot 72 h_{t-1} && \varepsilon_t \sim t(0, h_t, \lambda), \quad \lambda = 3 \cdot 45 \\
&\quad (1 \cdot 17) \quad (3 \cdot 3) \qquad (11 \cdot 7) && \qquad\qquad\qquad\qquad (4 \cdot 80)
\end{aligned}
\tag{30}
$$

Through the asymmetry of some of the marginal posteriors the 't-values' (or standard deviations of the posteriors) obtained from the Importance Sampling method are more reliable then those obtained by numerical differentation shown with the maximum likelihood estimates in equation (29).

In this stationarity analysis of the conditional variances of the US treasury bill rate we analyse only the parameters which concern the conditional variances, i.e. γ, α, β, and λ. Spillovers between these parameters and those of the level of the series, c (constant term) and ρ (AR(1)), are discussed in the following section. For the stationarity analysis we are not especially interested in the values of the parameters, per se but more in the values of certain functions of the parameters. The first function which we analyse indicates the 'finiteness' of the unconditional variance, $1 - \alpha E(z^2) - \beta = 1 - \alpha\lambda/(\lambda - 2) - \beta$ (see equation (18)). When this parameter function is negative, the unconditional variance is 'infinite'. In Figure 5 the contour lines of the bivariate posterior of $1 - \alpha\lambda/(\lambda - 2) - \beta$ and λ are shown. Given that the second-order moments of t-distributions with less then two degrees of freedom are infinite, the contour lines have a hyperbolic shape, indicating the 'infinity' of the means of the conditional posterior of $1 - \alpha\lambda/(\lambda - 2) - \beta$ given a value of λ, for $\lambda \leqslant 2$. This hyperbolic shape also shows the large positive correlation between λ and $1 - \alpha\lambda/(\lambda - 2) - \beta$, which is equal to $0 \cdot 63$. The mode of the bivariate posterior of $1 - \alpha\lambda/(\lambda - 2) - \beta$ and λ lies at a negative value of $1 - \alpha\lambda/(\lambda - 2) - \beta$. The 'infinity' of the unconditional variance and consequently the weak non-stationarity of the conditional variances are again quite unusual viewed from the perspective of the realized values of the conditional variances and the probabilities of increases or decreases in the conditional
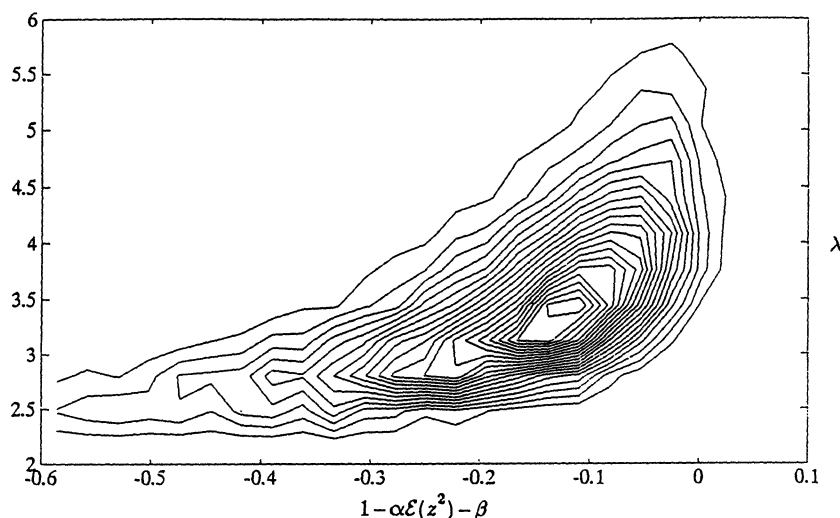
Figure 5. Contour lines: bivariate posterior $1 - \alpha E(z^2) - \beta = 1 - \alpha\lambda/(\lambda - 2) - \beta$ and λ

Moments posteriors	Mean	Std dev.
$1 - \alpha\lambda/(\lambda - 2) - \beta$	$-0\cdot17$	$0\cdot12$
λ	$3\cdot45$	$0\cdot72$
$\mathrm{corr}(1 - \alpha\lambda/(\lambda - 2) - \beta, \lambda)$	$0\cdot63$	

variances shown in Figures 2 and 3. Again the weak stationarity concept is quite misleading in the sense that the posteriors show that there is almost no probability for the stationary alternative, while the realizations (conditional variances, probabilities) are quite indecisive about stationarity or non-stationarity.

The marginal posterior of $1 - \alpha\lambda/(\lambda - 2) - \beta$ (see Figure 8) shows the infinity of the unconditional variance. This infinity automatically implies that the conditional variances are weakly non-stationary. In the previous section we have, however, shown the inappropriateness of the weak stationarity concept in the GARCH(1, 1) model and derived a quasi-strict stationarity measure. We would therefore like to analyse the posteriors involving this latter measure, $E(\ln(\beta + \alpha z^2))$.

While one can still calculate likelihood ratio statistics to test the hypothesis of weakly non-stationary conditional variances through the linear restrictions it imposes on the parameters, it will be very hard to construct a LR statistic to test the hypothesis of non-stationary conditional variances using the proper criterion $E(\ln(\beta + \alpha z^2))$. A procedure is to approximate the likelihood in the ML parameter point by some probability density function and to evaluate the value of $E(\ln((\beta + \alpha z^2))$ for parameter values generated from this function. This procedure accounts for the parameter uncertainty while a method based on the sample values of $\ln(\beta + \alpha z_t^2)$, $z_t = \varepsilon_t / h_t^{1/2}$, $t = 1, \ldots, T$, does not incorporate parameter uncertainty. The asymmetry of the likelihood, however, complicates the construction of a proper approximating function. A Bayesian kind of analysis which can properly deal with the asymmetry of the likelihood is therefore preferred.

To construct the marginal posterior of $E(\ln(\beta + \alpha z^2))$, we simulate sequences z_i, $i = 1, \ldots, n$, from a Student t-distribution with a degrees-of-freedom parameter resulting from the

generated parameter value from the Importance Function. For the generated parameter value from the Importance Function we approximate $E(\ln(\beta + \alpha z^2))$ by $n^{-1} \sum_{i=1}^{n} \ln(\beta + \alpha z_i^2)$, where both α and β are generated by the Importance Function. The marginal posterior of $E(\ln(\beta + \alpha z^2))$ is then calculated by using the weights attached to each drawing of $E(\ln(\beta + \alpha z^2))$, resulting from the Importance Sampling procedure. For more details on Importance Sampling, see Hop and van Dijk (1992).

It is interesting to analyse the relationship between the value of $E(\ln(\beta + \alpha z^2))$ and the degrees-of-freedom λ or the measure for weak stationarity ('finite' unconditional variance), $1 - \alpha E(z^2) - \beta = 1 - \alpha \lambda/(\lambda - 2) - \beta$. First, in Figure 6 the contour lines of the bivariate posterior of $E(\ln(\beta + \alpha z^2))$ and λ are shown. As can be concluded from these contour lines, the correlation between the degrees of freedom and the theoretical stationarity measure is negative (-0.21). Imposing normality therefore leads to more stationary conditional variances.

Second, Figure 7 contains the contour lines of the bivariate posterior of $E(\ln(\beta + \alpha z^2))$ and the 'finite' unconditional variance measure $1 - \alpha E(z^2) - \beta = 1 - \alpha \lambda/(\lambda - 2) - \beta$. The contour lines show that the two measures are negatively correlated, with a correlation coefficient equal to -0.38. This negative correlation arises because of the first-order Taylor expansion, $E(\ln(\beta + \alpha z^2)) = -(1 - \beta - \alpha E(z^2)) + \text{error}$ (see Section 2). The negative correlation is therefore quite natural. Also, given the fact that $E(\ln(\beta + \alpha z^2))$ should be positive to achieve non-stationarity while $1 - \beta - \alpha E(z^2)$ should be negative to have an 'infinite' unconditional variance. The small value of the correlation coefficient indicates, however, that the error term in the Taylor expansion is quite large, implying distinct differences between the strict measure and the 'finite' unconditional variance measure. Approximating the perhaps difficult to calculate measure $E(\ln(\beta + \alpha z^2))$ by $1 - \beta - \alpha E(z^2)$ leads to rather large errors as a consequence. To compare the two measures a little further, the marginal posteriors of the two different measures, $E(\ln(\beta + \alpha z^2))$ and $1 - \alpha E(z^2) - \beta$, are shown in Figure 8. These posteriors show that the probability of non-stationary conditional variances differs considerably between the two measures. The 'finite' unconditional variance measure,
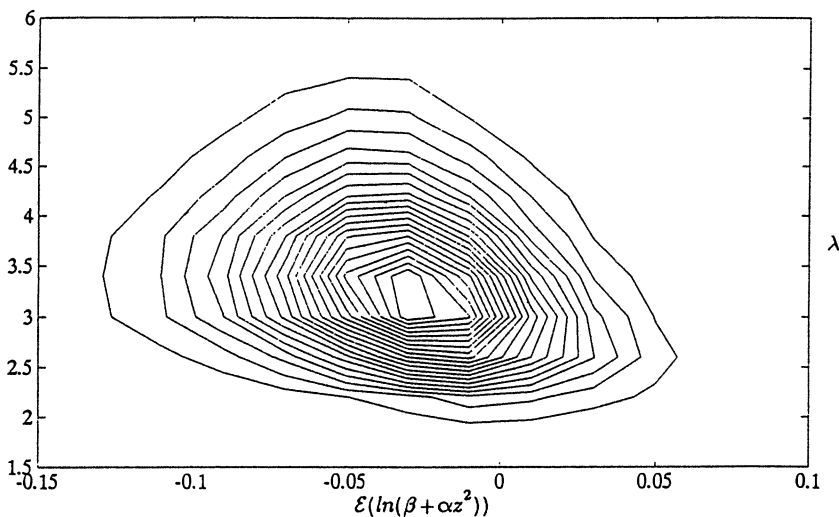

Figure 6. Contour lines: bivariate posterior $E(\ln(\beta + \alpha z^2))$ and λ
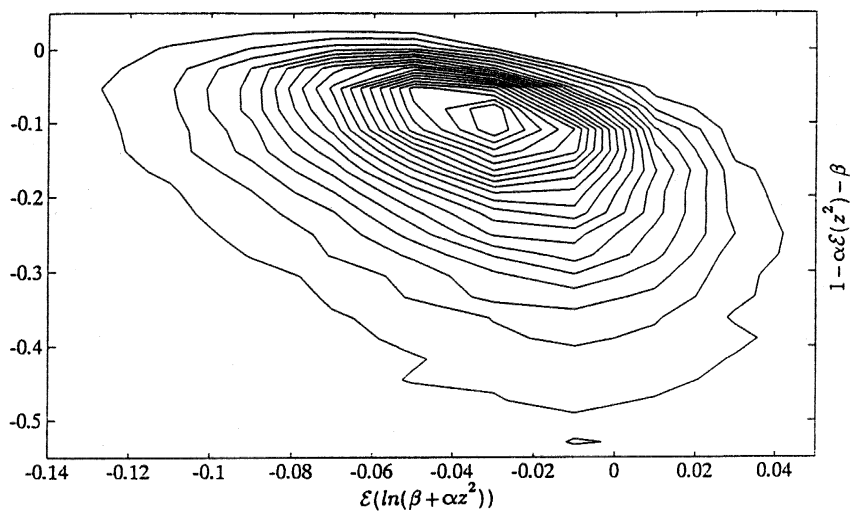
Figure 7. Contour lines: bivariate posterior $E(\ln(\beta + \alpha z^2))$ and $1 - \alpha E(z^2) - \beta = 1 - \alpha\lambda/(\lambda - 2) - \beta$

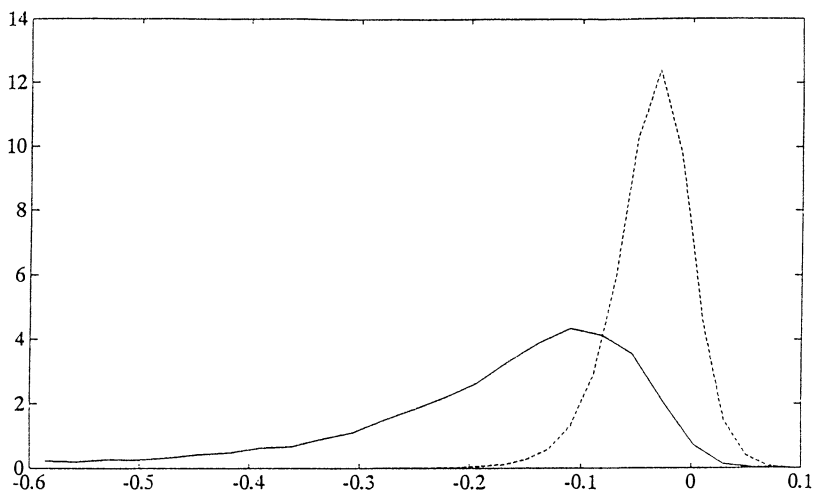Figure 8. Marginal posteriors $E(\ln(\beta + \alpha z^2))$ $(- -)$ and $1 - \alpha E(z^2) - \beta = 1 - \alpha\lambda/(\lambda - 2) - \beta$ (---)

Moments posteriors	Mean	Std dev.
$E(\ln(\beta + \alpha z^2))$	$-0\cdot036$	$0\cdot035$
λ	$3\cdot45$	$0\cdot72$
$1 - \alpha\lambda/(\lambda - 2) - \beta$	$-0\cdot17$	$0\cdot12$
$\text{corr}(E(\ln(\beta + \alpha z^2)), \lambda)$	$-0\cdot17$	
$\text{corr}(E(\ln(\beta + \alpha z^2)), 1 - \alpha\lambda/(\lambda - 2) - \beta)$	$-0\cdot38$	

$1 - \alpha E(z^2) - \beta$, leads to a probability of non-stationarity of approximately $0\cdot99$ while the probability of non-stationarity using the proper measure $E(\ln(\beta + \alpha z^2))$ is only $0\cdot13$. These two probabilities also allow one to approximate the posterior odds ratio. To calculate this ratio one has to specify prior odds $(1\cdot0)$ and a prior for the different hypotheses. To avoid the Lindley paradox we specified flat priors on the 99% Highest Posterior Density regions of the parameters under the different hypotheses. The posterior odds are then approximately equal to the ratio of the products of the prior and the posterior probabilities. For the parameters of interest, $E(\ln(\beta + \alpha z^2))$ and $1 - \beta - \alpha E(z^2)$, the posterior odds then become

$$H_0: E(\ln(\beta + \alpha z^2)) \geqslant 0, \qquad H_1: E(\ln(\beta + \alpha z^2)) < 0$$
$$K_{01} \qquad (0\cdot13/0\cdot07)/(0\cdot77/0\cdot18) = 0\cdot434$$
$$H_0: 1 - \beta - \alpha E(z^2) \leqslant 0, \qquad H_1: 1 - \beta - \alpha E(z^2) > 0$$
$$K_{01} \qquad (0\cdot99/0\cdot6)/(0\cdot01/0\cdot03) = 4\cdot95$$

The two probabilities and, correspondingly, the two posterior odds ratios differ substantially. Using the 'finite' unconditional variance measure, stationarity is rejected while the proper measure rejects non-stationarity. In the articles referred to by Bollerslev *et al.* (1992) the possible non-stationarity of the conditional variances of interest rates was always tested using the 'finite' unconditional variance measure, $1 - \alpha E(z^2) - \beta$, which equals $1 - \alpha - \beta$ in the case of normally distributed disturbances. Thus we have shown (1) that $1 - \alpha E(z^2) - \beta$ is a theoretically wrong condition for stationarity and (2) that it has also only a small correlation with the proper condition such that when using it as a non-stationarity condition, the probability of non-stationarity will be overestimated.

The importance of the less restrictive non-stationarity conditions using $E(\ln(\beta + \alpha z^2))$ instead of $1 - \alpha E(z^2) - \beta = 1 - \alpha \lambda/(\lambda - 2) - \beta$ lie especially in the area of forecasting. Using the standard forecasting formulae of the conditional variances (see Bollerslev, 1986), the forecasts are going to converge to infinity when $1 - \alpha E(z^2) - \beta$ is below 0. The non-stationarity measure $E(\ln(\beta + \alpha z^2))$ shows, however, that even for values of $1 - \alpha E(z^2) - \beta$ below 0, a finite limiting conditional variance is still possible. It is much better therefore to base the forecasts on direct simulations from models (9)–(11) than to use the forecasting formulas.

For the analysed treasury bill rate is not only the non-stationarity in the conditional variances but also that in the level of importance. The possible non-stationarity in the level is discussed in the next section.

4. SENSITIVITY OF THE UNIT ROOT WITH RESPECT TO FAT-TAILED DISTURBANCE DISTRIBUTIONS

In Section 3 we have shown that the hypothesis of a unit root in the AR(1) polynomial cannot be rejected for the US 3-month treasury bill rate series given the values of the ML estimate of the AR(1) parameter, ρ, and the mean of the posterior (both were $1\cdot0$). A ML estimate or posterior mean of the parameter ρ exactly equal to $1\cdot0$ is rather odd, however, and is due partly to the assumption of t-disturbances. For normal distributed errors (t-distributed with infinite degrees of freedom), the ML estimates and posterior means lie well below $1\cdot0$ and this indicates that the parameter ρ is negatively correlated with the degrees-of-freedom parameter of the t-distribution of the disturbances. In Figure 9 the marginal posterior using flat priors of the parameter ρ for four different AR(1) models for the treasury bill rate are shown. These models are (1) constant variance, normal distributed errors, (2) GARCH(1, 1) conditional

variances, normal distributed errors, (3) constant variance, t-distributed errors, and (4) GARCH(1, 1) conditional variances, t-distributed errors. Figures 10 and 11 contain the contour lines of the bivariate posteriors of ρ and the degrees-of-freedom parameter λ for the constant variance model (10) and the model with GARCH(1, 1) conditional variances (11). The contour lines in this latter figure clearly show the negative correlation $(-0\cdot14)$ between the degrees-of-freedom parameter λ and the AR parameter ρ. The negative correlation for the constant variance model is considerably smaller $(-0\cdot046)$ but still negative. Another interesting feature which shows the negative correlation between the degrees of freedom and the AR parameter
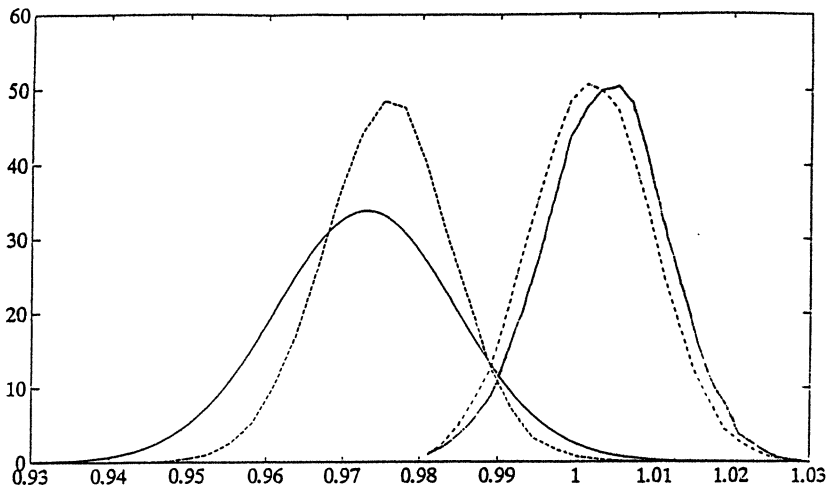

Figure 9. Marginal posteriors ρ, cons. var., norm. (—), GARCH(1, 1), norm. (– –), cons. var., t ($\cdots$), GARCH(1, 1), t ($\cdot - \cdot$)
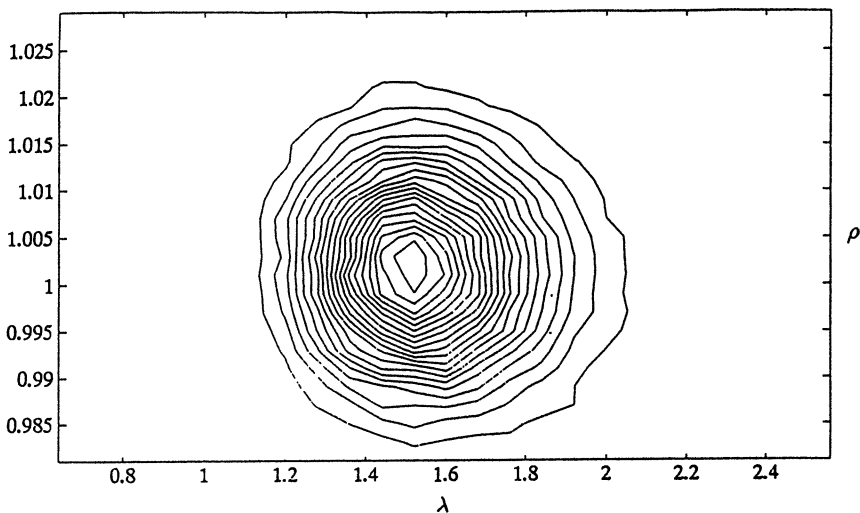

Figure 10. Contour lines: bivariate posterior ρ and λ, constant variance

ρ is the slight increase in the mode of the marginal posterior of ρ by imposing GARCH(1, 1) conditional variances when the disturbances are normally distributed. A GARCH(1, 1) model leads namely to a larger kurtosis of the disturbances and can therefore be compared with a certain kind of t-disturbances with finite degrees of freedom. It is also interesting to note that imposition of constant variances leads to a higher variance of the marginal posterior of ρ when the disturbances are normally distributed. This is similar to the well-known property that the OLS estimator is consistent but not efficient in the presence of heteroscedasticity. The increase in the variance does not occur when the disturbances are t-distributed because with a lower value of the parameter λ one can also explain occurrence of several rather large values of the disturbances, i.e. heteroscedasticity. The parameter λ is therefore a good approximation of the degree up to which one is able to explain the variance/kurtosis of the disturbances. Models for the conditional variances which do not explain the variance/kurtosis properly will lead to lower values of the parameter λ than models which explain a larger part of the variance/kurtosis. The treasury bill series confirm these statements, as can be concluded from Figure 12, where the two marginal posteriors of λ are drawn for a model with GARCH conditional variances and constant variances. Figure 12 shows that imposition of a constant variance leads to t-distributed disturbances which degrees-of-freedom parameter is almost always below $2 \cdot 0$. The variance of the disturbances is infinite as a consequence. The posterior of λ for GARCH(1, 1) conditional variances lies almost completely above $2 \cdot 0$ such that the conditional variances of the disturbances are finite.

Figure 9, where the marginal posteriors of the parameter ρ are shown, allows us to test for the presence of unit roots. One straightforward method for testing the unit root hypothesis is to use the Highest Posterior Density regions (HPD). Figure 9 shows that $1 \cdot 0$ lies in the 95% HPD regions of the models having t-distributed disturbances and the presence of $1 \cdot 0$ in the 95% HPD regions of the models with normal distributed disturbances is doubtful. Given the marginal posteriors in Figure 9, it is also possible to approximate the posterior odds of the unit root hypothesis, $H_0: \rho = 1$, against $H_1: \rho \neq 1$. Using a flat prior over the 99% HPD region of ρ to avoid the Lindley paradox and a prior odds ratio equal to $1 \cdot 0$, the value of the posterior
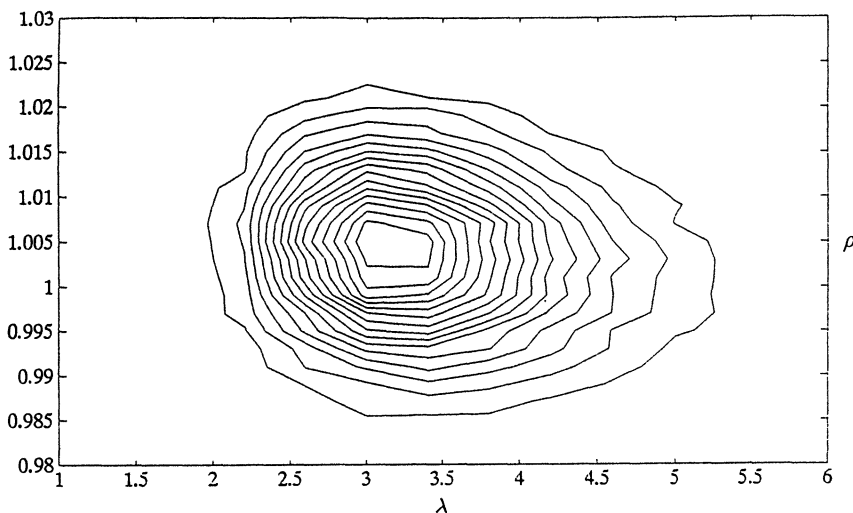

Figure 11. Contour lines: bivariate posterior ρ and λ, GARCH(1, 1) contour variances
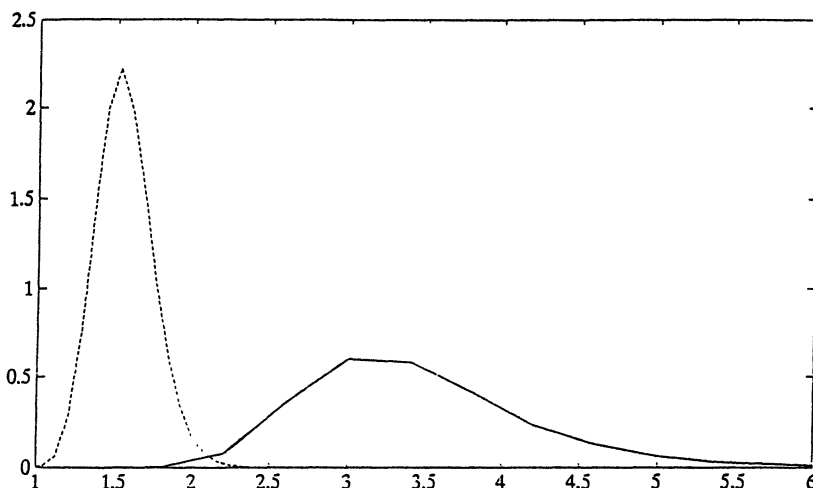
Figure 12. Marginal posterios λ, GARCH(1, 1) conditional variances (—) and constant variance (– –)

Moments posteriors	Mean	Std dev.
ρ cons. var., norm.	0·971	0·012
GARCH(1, 1), norm.	0·983	0·0078
cons. var., t	1·002	0·0077
GARCH(1, 1), t	1·004	0·0077
λ cons. var.	1·57	0·26
GARCH(1, 1)	3·45	0·72
corr($\rho_{cons.,t}$, λ)	−0·046	
corr($\rho_{GARCH.,t}$, λ)	−0·14	

odds equals approximately the product of the length of the 99% HPD region and the value of the marginal posterior in 1·0. It has to be mentioned that the posterior odds ratio calculated in this way does not test the proper hypothesis H_0: $\rho = 1$ and $c = 0$ (constant term), which affects two parameters instead of one. Calculation of the posterior odds using only the parameter ρ to test the hypothesis of a unit root can be justified using the argument that when $\rho = 0$, c automatically becomes 0 because there is no linear time trend in the treasury bill rate. The 't-value' of the posterior mean of c (0·68) also indicates that 0 lies in 95% HPD region of the marginal posterior of c, which corresponds to the absence of a linear time trend. In classical statistics one is also allowed to use only the parameter ρ to test for unit roots when there is no linear time trend in the analysed data series. The posterior odds for the hypothesis of a unit root in the four different AR(1) models then become

K_{01}	cons. var., norm.	0·17
	GARCH(1, 1), norm.	0·04
	cons. var., t	2·30
	GARCH(1, 1), t	2·50

Using the formulas from Geweke (1989a), it is possible to calculate the different posterior odds accurately, including those for the proper hypothesis H_0: $\rho = 1$ and $c = 0$. The posterior odds mentioned above test for a unit root in the AR(1) polynomial. The unit root is, however, only one of the possible non-stationary parameter values, and tests for non-stationarity can also be performed using the hypothesis, H_0: $\rho \geqslant 1$ and H_1: $\rho < 1$.

The sensitivity of the AR parameter ρ with respect to the value of the degrees-of-freedom parameter λ is remarkable and can be explained using robustness arguments. Using constant variances and normal distributed disturbances one mainly focuses on proper modelling of the observations having relatively large disturbances. Minus the derivative of the log of the Probability Density Function (PDF) of the disturbances with respect to the squared random variable is namely a constant function of the random variable in the case of a standard normal PDF and a decreasing function in the case of a t PDF:

$$p_{\text{normal}}(\varepsilon) = (2\pi)^{-1/2} \exp\left[-\tfrac{1}{2}\varepsilon^2\right] \Rightarrow -\frac{\partial \ln p_n}{\partial \varepsilon^2} = \tfrac{1}{2}$$

$$p_t(\varepsilon) = \frac{\Gamma(\tfrac{1}{2}(\lambda+1))}{\Gamma(\tfrac{1}{2}\lambda)\Gamma(\tfrac{1}{2})} \lambda^{-1/2} \left|1 + \frac{\varepsilon^2}{\lambda}\right|^{-1/2(\lambda+1)} \Rightarrow -\frac{\partial \ln p_t}{\partial \varepsilon^2} = \frac{1}{2}\frac{(\lambda+1)}{(\lambda+\varepsilon^2)}$$

(31)

In Figure 13 minus the derivatives of the log of three different PDFs, i.e. normal, t with one degree of freedom (Cauchy) and t with 10 degrees of freedom are shown. While the derivatives for the normal remain constant over different values of the random variable, ε, the derivatives decline for increasing values of the absolute value of the random variable, ε, in the case of the t PDFs. For decreasing values of the degrees-of-freedom parameter λ, this declining behaviour becomes even more pronounced. As a consequence, a log likelihood consisting of the sum of T of such log PDFs will have a constant sensitivity with respect to an increase or decrease in a certain disturbance in case of normal distributed disturbances, while the sensitivity differs over the observations in the case of t-distributed disturbances, being larger for relatively small values of the disturbances. The declining sensitivity of the log likelihood for increasing values of the disturbances (outliers) leads to an interesting aspect of modelling using t-distributed disturbances, the resulting robust parameter estimation procedure with respect to outliers. The decreasing derivatives indicate that the resulting estimation procedure attaches less weight to relatively large disturbances (outliers). Estimation using normal distributed disturbances and constant variances attaches the same weight to all disturbances, and the resulting estimation procedure focuses as a consequence on the relatively large values
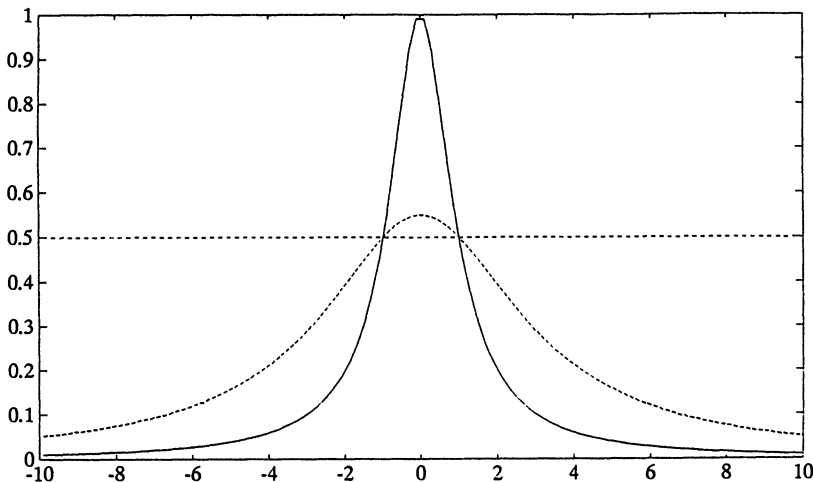

Figure 13. Minus derivative log PDF with respect to squared random variable, Cauchy (—), normal ($\cdot - \cdot$), t with $\lambda = 10$ (— —)

of the disturbances. The focus of the estimation procedure on the large values of the disturbances (outliers) declines when the degrees-of-freedom parameter of the *t*-distribution decreases. Estimation using *t*-disturbances leads therefore to more robust ML estimators, posterior means with respect outliers.

In the analysed data series such outliers are clearly present (see Figures 1 and 2) and possibly occur during the years 1975 and 1980–84. The arguments stated above show that when using normal distributed disturbances one especially focuses on these years of outlying observations and a large weight is implicitly attached to these observations. Using GARCH conditional variances (small decrease in weight) and/or *t*-distributed disturbances these weights are decreased and more even weights are attached to the different observations.

In the growing literature on Bayesian unit root analysis (see Schotman and van Dijk, 1991; Phillips, 1991; DeJong and Whiteman, 1991) one concentrates mainly on the issue of the construction of proper priors. The analysis of the treasury bill rate has been conducted using flat priors which favour stationarity in models having normal distributed disturbances compared with the priors used by Phillips and Schotman and van Dijk. The sensitivity of the unit root with respect to the degrees of freedom of the *t*-distribution of the disturbances raises a new issue of importance, the proper modelling of the distribution of the disturbances. Geweke (1992) showed that the 'unit roots' in the well-known Nelson–Plosser series seem to be rather insensitive to the degrees-of-freedom parameter λ. These series are not very heteroscedastic, however, and contain only a few outliers resulting from structural breaks. For the interest rate, and probably also for other financial series, the heteroscedasticity and presence of outliers are two of the important features of the series, and neglecting these properties, by imposing normality and/or GARCH/constant variances, can lead to quite different results (see, for instance, the treasury bill series).

5. CONCLUSIONS

Interpreting non-stationarity in terms of infinite persistence of shocks, we show that the non-stationarity properties of the conditional variances of GARCH(1, 1) models differ from the weak non-stationarity properties analysed mainly in the literature. Using the weak stationarity conditions, the literature shows, for example, that the US 3-month treasury bill rate has non-stationary conditional variances. Application of the quasi-strict stationarity conditions, defined in this paper and related to the concept of infinite persistence of shocks, shows, however, that the conditional variances of these series are likely to be stationary.

Also, interest rate series are likely to be non-stationary in the level of the series. Standard unit root analysis often assumes that the disturbances are normally distributed. The treasury bill rate behaves in quite a volatile manner and the assumption of normally distributed disturbances is therefore not quite realistic. Applying independent *t*-distributed disturbances gives a remarkable phenomenon, i.e. the negative correlation between the degrees-of-freedom parameter of the *t*-distributed disturbances and the unit root parameter. By relaxing the assumption of normally distributed (*t* with infinite degrees of freedom) disturbances, the probability of a unit root in the AR(1) model increased considerably. Especially for financial series, the sensitivity of the unit root with respect of the assumption of normality will be important due to the volatile nature of these series. The importance of the proper construction of the probability density function of the disturbances raises a new issue in the discussion of Bayesian unit root analysis, which has been focused primarily on construction of priors so far. For some of the series analysed in this discussion (i.e. financial series) the assumption of normality can be quite restrictive and the results may be sensitive to this assumption.

REFERENCES

Bera, A. K., and M. L. Higgins (1992), 'A survey of ARCH models: properties, estimation and testing', *Journal of Economic Surveys*, forthcoming.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroscedasticity', *Journal of Econometrics*, **31**, 307–327.

Bollerslev, T., R. Y. Chou and K. F. Kroner (1992), 'ARCH modelling in finance: a selective review of the theory and empirical evidence', *Journal of Econometrics*, **52**, 5–59.

Box, G. E. P., and G. W. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

DeJong, D. N., and C. H. Whiteman (1991), 'The temporal stability of dividends and stock prices: evidence from the likelihood function', *American Economic Review*, **81**, No. 3, 600–617.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the United Kingdom inflation', *Econometrica*, **50**, 987–1008.

Engle, R. F., and T. Bollerslev (1986), 'Modelling the persistence of conditional variances', *Econometric Reviews*, **5**, 1–50.

Gelfand, A. E., and A. F. M. Smith (1990), 'Sampling based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398–409.

Geweke, J. (1986), 'Modelling the persistence of conditional variances: comment', *Econometric Reviews*, **5**, 57–61.

Geweke, J. (1989a), 'Exact predictive densities for linear models with ARCH disturbances', *Journal of Econometrics*, **40**, 63–86.

Geweke, J. (1989b), 'Bayesian inference in econometric models using Monte Carlo integration', *Econometrica*, **56**, 1317–1339.

Geweke, J. (1992), 'Bayesian treatment of the independent Student-*t* linear model', Working paper, Department of Economics, Federal Reserve bank of Minneapolis.

Hop, J. P., and H. K. van Dijk (1992), 'SISAM and MIXIN: two algorithms for the computation of posterior moments and densities using Monte-Carlo integration', *Computer Science in Economics and Management*, **5**, 183–220.

Kloek, T., and H. K. van Dijk (1978), 'Bayesian estimates of equation system parameters: an application of integration by Monte Carlo', *Econometrica*, **46**, 1–20.

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

Phillips, P. C. B. (1991), 'To criticize the critics: an objective analysis of stochastic trends', *Journal of Applied Econometrics*, **6**, 333–364.

Schotman, P., and H. K. van Dijk (1991), 'A Bayesian analysis of the unit root in real exchange rates', *Journal of Econometrics*, **49**, 195–238.

Spanos, A. (1991), 'On modeling heteroscedasticity: the Student's *t* and elliptical linear regression models', Working paper, Department of Economics, Virginia Polytechnic Institute.

Van Dijk, H. K., and T. Kloek (1980), 'Further experience in Bayesian analysis using Monte Carlo integration', *Journal of Econometrics*, **14**, 307–328.