Structural Change and long memory in the GARCH(1,1)-model [1]

by

Walter Krämer and Baudouin Tameze Azamo
Fachbereich Statistik, Universität Dortmund, Germany
Phone xx231/755-3125, Fax: xx231/755-5284
e-mail: walterk@statistik.uni-dortmund.de

version October 2006

Summary

It has long been known that the estimated persistence parameter in the GARCH(1,1) - model is biased upwards when the parameters of the model are not constant throughout the sample. The present paper explains the mechanics of this behavior for a particular class of estimates of the model parameters. It gives sufficient conditions for the estimated persistence to tend to one when the mean of the process changes, both for a given sample size (as the size of the structural change increases), and as sample size increases, extending previous results that were concerned with changes in the volatility parameters.

Keywords: structural change, long memory, GARCH.

[1]Research supported by Ruhr Graduate School of Economics, Dortmund-Bochum-Essen, and by Deutsche Forschungsgemeinschaft under SFB 475.

1 Introduction

The GARCH(1,1) - model,

$$\begin{aligned} x_t &= \epsilon_t + \mu \\ \epsilon_t &= \eta_t \sigma_t \\ \sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned} \tag{1}$$

where $\eta_t \sim iid(0,1)$ and η_t is independent of the past ϵ's and σ's, is still the main workhorse in all areas of applied economics whenever conditional heteroscedasticity among the x_t is important. Typical examples include stock returns or inflation rates. Almost from the moment it was born, it was however plagued by the observation that in many applications, the estimate of the "persistence parameter" $\delta := \alpha + \beta$, no matter in which way obtained, was viewed as much too large (in the sense that the superior forecasting performance implied by high persistence did not materialize in empirical applications), and that this upward bias towards the maximum of 1 increases with increasing sample size.

For illustration, figure 1 plots various estimates that have been reported in the literature against the sizes of the respective samples. The number attached to the data points are sample sizes; they may be use to identify the papers. For ease of comparison, we confine ourselves to studies which use daily data (either FX-returns or stock returns; a more detailed description can be found in the appendix). The figure clearly demonstrates that estimated persistence increases with sample size and is almost indistinguishable from unity for samples of size 2000 or more.

Focusing on daily data ensures that sample size is proportional to calendar time, which appears to be the real driving force behind the increase in the estimated persistence. With hourly data, and a sample size of about 3000, Baillie and Bollerslev (1990) obtain estimates of persistence only in the range 0.4 - 0.7, while with monthly data, the estimated persistence is already above 0.9 for sample sizes around 500. Therefore the upward tendency in the estimated persistence is due to an increase in calendar time, not to an increase in sample size as such.

Diebold (1986) was probably the first to point out that this upward tendency of estimated $\delta's$ might be due to a switch in regime somewhere in the sample, the probability of which increases with increasing calendar time. Among many others Lamoreux et al. (1990), Hamilton and Susmel (1994) or Mikosch and Starica (2004), show that empirical estimates of δ do indeed decrease when the sample is split according to some sensible criterion, and they propose generalizations of (1) to account for changes in the parameters.

When standard GARCH(1,1)-models are fitted to data generated from such more general models, empirical estimates $\hat{\delta}$ of δ are rather close to, but usually less than one. Haas et al. (2004) figure 1 show by Monte Carlo simulations that $\hat{\delta}$ approaches 1 as persistence in their Markov - switching model increases; Mikosch and Starica (2004) show analytically that the Whittle - estimator of δ becomes arbitrarily close to 1 if the differences in the variances of their sub-models tend to infinity, and Hillebrand (2005) proves the same for ML-estimators for the case when the number of structural changes remains finite as sample size increases. The present paper considers the Minimum Distance Estimator (MDE) of α and β suggested by Baillie and Chung (2001), and shows that the sum of the estimated α and β can likewise be made arbitrarily close to 1 if there are structural changes in the unconditional expectation μ of the x_t-process, or more generally, if the ${x_t}^2$-process behaves as if it had nonstationary long memory.

2 Structural change in the mean and sample correlations

The point of departure of this paper is the relationship between certain types of structural change in the model (1) and the estimated autocorrelations of the ϵ_t^2. Most models that allow for changes in the coefficients of (1) do so by letting μ, ω, α or β depend on the (unobserved) state of a finite - dimensional Markov chain. Recent examples and variants thereof, with useful surveys of the literature, are Klaassen (2002) or Haas et al. (2004). Alternatively, Hamilton and Susmel (1994) or Wong and Li (2001) consider

$$\epsilon_t^* := g(\Delta_t)\epsilon_t, \tag{2}$$

where ϵ_t is generated by (1) (or some variant thereof), and g again depends on the state of some Markov-process $\{\Delta_t\}$ or some other stochastic process. Here, structural changes do not affect the dynamics of the process, just the scale. Other examples are Dueker (1997), who considers changes in the variance of the innovations η_t, or Mikosch and Starica (2004) and Hillebrand (2005), who simply collect together different sub-samples from different stationary models. All of these models imply that $E(x_t^2)$ is not constant over time.

The present paper considers the Minimum-Distance estimator of α and β when there are structural changes in the unconditional expectation μ which are ignored when the model (1) is fitted to the data. These changes can be both deterministic or stochastic, for instance, by letting μ depend on the state of an independent Markov process Δ_t:

$$x_t := \mu(\Delta_t) + \epsilon_t. \tag{3}$$

This is similar to (2), except that it is the conditional mean and not the conditional variance of x_t that is affected. No matter which way the process changes, it is easily seen that any such change will in general increase the empirical autocorrelations of the x_t^2.

For structural changes in μ, other than for structural changes in ω, α and β, there will, in addition to an increase in the empirical autocorrelations of the x_t^2's, also be an increase in the empirical autocorrelations of the x_t's themselves. This holds for all types of stochastic processes, not just GARCH(1,1). For illustration, figure 2 depicts the first 16 empirical autocorrelations computed from $n = 4000$ observations, for a stationary MA(2) process

$$x_t = \mu + \epsilon_t + 0.5\epsilon_{t-1} + 0.5\epsilon_{t-2}, \epsilon_t \sim nid(0,1),$$

where μ switches from -d to d in the middle of the sample. Without such change in μ, the theoretical autocorrelations are $\rho_1 = 0.5$, $\rho_2 = 0.17$, $\rho_3 = \rho_4 = ... = 0$. As the figure shows, estimated correlations are much larger and tend to 1 as d increases.

Let in general x_t $(t = 1, ..., T)$ be any short memory sequence of random variables with bounded variance and k shifts in mean at $1 < t_1 < ... < t_k < T$, and consider the empirical h'th order autocorrelation coefficient

$$\hat{\rho}_h = \frac{\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \ .$$

Rewriting the numerator as

$$\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x}) = \sum_{t=1}^{T}(x_t - \bar{x})^2 - \sum_{t=T-h+1}^{T}(x_t - \bar{x})^2 + \sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - x_t), \tag{4}$$

we see that

$$\hat{\rho}_h = 1 - \frac{\sum_{t=T-h+1}^{T}(x_t - \bar{x})^2}{\sum_{t=1}^{T}(x_t - \bar{x})^2} + \frac{\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - x_t)}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \ , \tag{5}$$

where the last two terms can be made as close to 0 as desired if n is "large" relative to k and $\sum(x_t - \bar{x})^2 \xrightarrow{P} \infty$. This is so because the first term tends to zero as $h/n \to 0$, and the second term tends to zero in view of the fact that $(x_t - \bar{x})(x_{t+h} - x_t)$ is "small" relative to $(x_t - \bar{x})^2$ whenever x_{t+h} and x_t belong to the same regime. When the number of shifts is small relative to sample size, this will apply to an increasing number of terms in the sum, so the ratio becomes arbitrarily small and empirical autocorrelation of th $x_t's$ can be made as close to unity as desired for any given sample size.

Another avenue through which empirical autocorrelations may be led to tend to 1 is for increasing sample size, when the x_t can be made to behave as if they were I(d) with $d \geq \frac{1}{2}$:

$$Var\left(\sum_{t=1}^{T} x_t \right) \ = \ O\left(\ T^{2d+1} \ \right). \tag{6}$$

It has long been known (see e.g. Krämer (1985)) that for $d = 1$, empirical autocorrelations of x_t of all orders must tend to 1 in probability as $T \longrightarrow \infty$, and Hassler (1997) shows that this holds for fractional integration parameters with $\frac{1}{2} \leq d < 1$ as well. The intuition behind this is that the last two terms in expression (5) become arbitrarily small as $T \longrightarrow \infty$ as the numerators are of smaller orders in probability than the denominators.

Diebold and Inoue (2001) show that behavior of type (6) occurs for instance whenever μ is stochastic and independent of ϵ_t and displays structural breaks of the form

$$\mu_t = \mu_{t-1} + \nu_t \tag{7}$$

$$\nu_t = \begin{cases} 0 & \text{with probability} \quad 1-p \\ \omega_t & \text{with probability} \quad p, \end{cases}$$

where $\omega_t = i.i.d(0, \sigma^2)$, and where p may depend on sample size. Since

$$\sum_{t=1}^{T} \mu_t = T v_1 + (T-1) v_2 + ... + v_T, \tag{8}$$

we have

$$Var\big(\sum_{t=1}^{T} \mu_t\big) = p\ \sigma^2 \sum_{t=1}^{T} t^2 = p\ \sigma^2 \frac{T(T+1)(2T+1)}{6}, \tag{9}$$

so we can have (6) for any $d,\ 0 < d \leq 1$, by letting

$$p = c\frac{1}{T^{2-2d}} \quad (0 < c \leq 1). \tag{10}$$

Of course, in the limiting case where $d = 1$ and p does not depend on T, μ_t and therefore also x_t will be I(1) and long memory will be extreme.

Spurious long memory in x_t can also be induced by time varying staying probabilities in the Markov-switching model of (3). For two states and serially independent ϵ's we have:

$$\begin{aligned} p_{00} &= 1 - c_0 T^{-\delta_0} \\ p_{11} &= 1 - c_1 T^{-\delta_1} \end{aligned}$$

and Diebold and Inoue (2001) show that then (6) applies with

$$d = \frac{1}{2} max\big\{min(\delta_0, \delta_1) - |\delta_0 - \delta_1|, 0\big\}. \tag{11}$$

To the extent that this carries over to the case where the ϵ_t's follow a GARCH-process we will for $d_0 = d_1 = 1$ again have empirical autocorrelations of the x_t which tend to 1 as a consequence of structural change.

We will not enter into a detailed discussion of this phenomenon here. There might well be many other instances where this tendency towards unity of empirical autocorrelations occurs. Diebold and Inoue (2001) for instance show that the Engle and Smith (1999)–STOP-BREAK model, which generates an I(1)-series, can be generalized to an arbitrary I(d)-behavior where in all cases we have autocorrelations increasing with sample size. For the present purpose, it suffices to know that there do exist meaningful models which induce empirical autocorrelations of a time series to become large. The conditions that guarantee this to happen do not concern us here. Rather, we take this behavior as given and explore its implications for the estimated persistence of a GARCH(1,1)-model.

To that purpose, it remains to show that real or spurious long memory in the x_t's induces real or spurious long memory in x_t^2 (since the estimator which we consider in section 3 is based on the empirical autocorrelations of the squared observations). For a given sample size and increasing breaks, it is easily seen that the arguments that lead to increasing autocorrelations of x_t also lead to increasing autocorrelations of x_t^2. For "genuine" Gaussian I(d)-processes with $d \geq \frac{1}{2}$, Dittmann and Granger (2002) show that the squared process is also I(d) with the same d, and similar results hold for spurious long memory as well (in the sense that convergence to 1 of the empirical autocorrelations of the x_t's implies convergence to 1 of the empirical autocorrelations of the x_t^2's). For instance, it is easily seen that with μ's changing according to (7), the empirical autocorrelation of both the x_t's and the x_t^2's must tend to 1 as sample size increases.

3 Estimating persistence

Next we consider a particular estimator, the Baillie and Chung (2001)–Minimum-Distance-Estimator of α and β, given that empirical autocorrelations behave as explained in section 2. This estimator is based on the ARMA(1,1) -representation of ϵ_t^2 given by

$$\epsilon_t^2 = \omega + (\alpha + \beta)\epsilon_{t-1}^2 + u_t - \beta u_{t-1}, \tag{12}$$

where

$$u_t := \epsilon_t^2 - E(\epsilon_t^2 | \epsilon_{t-1}^2, \epsilon_{t-2}^2, ...) = \epsilon_t^2 - \sigma_t^2 \tag{13}$$

is white noise and uncorrelated with past ϵ_t^2's. We also require $\alpha > 0, \quad \beta \geq 0$ and $\alpha + \beta < 1$. The basic idea is to exploit the fact that, because of (12), the theoretical autocorrelations of ϵ_t^2 are known functions of α and β:

$$\begin{aligned}
\rho_1 &= \alpha + \frac{\alpha^2\beta}{1 - 2\alpha\beta - \beta^2} \\
\rho_2 &= (\alpha + \frac{\alpha^2\beta}{1 - 2\alpha\beta - \beta^2})(\alpha + \beta) \\
&\vdots \\
\rho_h &= (\alpha + \frac{\alpha^2\beta}{1 - 2\alpha\beta - \beta^2})(\alpha + \beta)^{h-1} \quad (h > 1).
\end{aligned} \tag{14}$$

The ρ_k are then estimated by

$$\hat{\rho_k} = \frac{\sum_{t=1}^{T-k}(\tilde{\epsilon}_t^2 - \bar{\tilde{\epsilon}}^2)(\tilde{\epsilon}_{i+k}^2 - \bar{\tilde{\epsilon}}^2)}{\sum_{t=1}^{T}(\tilde{\epsilon}_t^2 - \bar{\tilde{\epsilon}}^2)^2},$$

where $\tilde{\epsilon}_t := x_t - \bar{x}$, and the Minimum Distance Estimators $\hat{\alpha}$ and $\hat{\beta}$ for α and β are obtained as

$$\arg\min_{\alpha,\beta} [\hat{\rho} - \rho(\alpha, \beta)]' W [\hat{\rho} - \rho(\alpha, \beta)], \tag{15}$$

where W is some suitable positive definite weighting matrix, $\hat{\rho} = (\hat{\rho_1}, \ldots, \hat{\rho_h})'$ and where $\rho(\alpha, \beta) = (\rho_1, \ldots, \rho_h)'$ is a vector-valued function of α and β defined in (14).

The efficiency of this estimator relative to the Maximum Likelihood estimator is evaluated in detail by Baillie and Chung (2001); it depends on the particular choice of h and W and does not concern us here. Rather, we take h and W as given and consider the behavior of $\hat{\delta} = \hat{\alpha} + \hat{\beta}$ as

$$\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_h)' \xrightarrow{p} e := (1, \ldots, 1)'. \tag{16}$$

This particular limiting behavior of $\hat{\rho}$ implies that

$$\begin{aligned} &\arg\min_{\alpha,\beta} [plim\ \hat{\rho} - \rho(\alpha,\beta)]' W [plim\ \hat{\rho} - \rho(\alpha,\beta)] \\ &\quad \subseteq \arg\min_{\alpha,\beta} [e - \rho(\alpha,\beta)]' W [e - \rho(\alpha,\beta)], \end{aligned} \tag{17}$$

where the latter set of minimizing values of α and β is in view of (6) determined by

$$\alpha + \beta \quad = \quad 1 \qquad \text{and} \tag{18}$$

$$\alpha + \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2} \quad = \quad 1. \tag{19}$$

This is so because (18) and (19) are equivalent to $\rho(\alpha, \beta) = e$, which is equivalent to

$$[e - \rho(\alpha,\beta)]' W [e - \rho(\alpha,\beta)] = 0, \tag{20}$$

which in view of the positive definiteness of W is the minimum value which can be attained.

It is easily checked that (18) implies (19), so all pairs of α and β with $\alpha > 0$, $\beta \geq 0$ and $\alpha + \beta = 1$ are candidates for $plim_{\hat{\rho} \to e}(\hat{\alpha}, \hat{\beta})$. Which one of these will eventually materialize depends on the particular way in which $\hat{\rho}$ approaches e. In practice, it appears that small values of $\hat{\alpha}$ and large values of $\hat{\beta}$ are preferred (see e.g Haas et al. (2004), figure 1). The point of interest here is that no matter what the particular probability limits of $\hat{\alpha}$ and $\hat{\beta}$ are, they must always sum to 1.

Another line of reasoning, different from ours, which also leads to $\hat{\delta} \xrightarrow{p} 1$, is due to Hillebrand (2005): If the model (1) is estimated by Maximum Likelihood, we must have

$$\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha}\hat{\varepsilon}_{t-1}^2 + \hat{\beta}\hat{\sigma}_{t-1}^2 \quad (t = 1, \cdots, T), \tag{21}$$

where the $\hat{\sigma}_t^2$ and $\hat{\epsilon}_t$ are fitted values obtained from the ML-estimators $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$ and some starting values ϵ_0^2 and δ. If there are in addition finitely many regimes, with regime-specific stationary expectations $E(\sigma_t^2)_{(i)} = E(\varepsilon_t^2)_{(i)} =: E_i$ and with regime-specific sample sizes increasing, one obtains under certain conditions on the estimators that

$$\overline{\hat{\sigma}^2}_{(i)} \xrightarrow{p} E_{(i)}, \quad \overline{\hat{\varepsilon}^2}_{(i)} \xrightarrow{p} E_{(i)}, \tag{22}$$

so

$$E_{(i)} - E \approx (\hat{\alpha} + \hat{\beta})(E_{(i)} - E), \tag{23}$$

where E is the sample mean of the $\hat{\sigma}_t^2$, and therefore $\hat{\alpha}+\hat{\beta}$ must tend to 1. This argument however depends crucially on the validity of the limiting relationship in (22) and is different from the one advanced in the present paper.

4 Some finite sample simulations

This section summarizes various Monte Carlo simulations to check the finite sample relevance of the above results. First we consider the case where sample size is fixed and where there are increasing breaks in the mean μ of the x_t–series. Figure 3 shows the mean estimated persistence $\hat{\delta}$ (averaged over 1000 experiments) as a function of the size of the break in μ. The break is always in the middle of the sample, and we choose $\omega = 0.001$, $\alpha = 0.2$ and $\beta = 0.4$.

The figure shows that the estimated persistence rapidly approaches 1 as the size of the break increases. We do not show the estimated α's and β's separately, so the (almost) unbiased estimates of $\delta = \alpha+\beta$ in the absence of a break masks the well known fact that in correctly specified models, α is usually underestimated and β is usually overestimated in finite samples. Separate results for α and β are available from the authors upon request.

In all experiments, we use 10 lags for the minimum distance estimation and obtain the weighting matrix W via the Newey-West (1987)-procedure. Results however remain virtually unchanged for different number of lags and weighting schemes. Also, experiments with different α's and β's were performed, which all confirmed the message contained in figure 3. Figure 4 and 5 show results for $\alpha = \beta = 0.3$ and $\alpha = 0.4$, $\beta = 0.2$, respectively. Again, δ is estimated (almost) unbiasedly when there is no structural break, and the upward bias increases rapidly as the size of the break in μ increases.

In another series of experiments, we let μ change according to the Diebold and Inoue (2001)–scheme from equation (7). Figure 6 shows some sample time series of the μ's, and figure 7 reports the estimated persistence as a function of the sample size. As was to be expected, the upward bias again increases rapidly as sample size increases, irrespective of the switching probability.

5 Possible extensions

The arguments above extend naturally to more general GARCH(p, q)-models, where

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2.$$

The persistence is here measured by $\alpha_1 + \ldots + \alpha_p + \beta_1 + \ldots + \beta_q$ and it is easily seen that the Minimum Distance Estimator must likewise have the property that

$$\hat{\alpha_1} + \ldots + \hat{\alpha_p} + \hat{\beta}_1 + \ldots + \hat{\beta}_q \xrightarrow{p} 1$$

whenever $\hat{\rho}_i \xrightarrow{p} 1$ (i=1,2,3,...). This follows from the fact that theoretical autocorrelations ρ_i can be written as

$$\rho_i = g(\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)(\alpha_1 + \ldots + \alpha_p + \beta_1 + \ldots + \beta_q)^{i-1},$$

for some continuous function g, so for $\hat{\rho}_i \to 1$, the distance between empirical and theoretical autocorrelations is minimized for $\alpha_1 + \ldots + \alpha_p + \beta_1 + \ldots + \beta_q = 1$.

However, extensions to other types of structural shifts are not as obvious. For instance, if we also allow for shifts in ω, α or β, we have breaks in $E(\epsilon_t^2)$, but also large shifts in the variance of ϵ_t^2, so our argument leading to $\hat{\rho} \to 1$ for given sample size breaks down (in section 3, we had implicitly assumed that the variance remains bounded when there are structural breaks in the mean). One could still of course obtain persistence parameters close to unity, as $\hat{\rho}_h \xrightarrow{p} 1$ is only a sufficient, not a necessary condition for $\hat{\alpha} + \hat{\beta} \xrightarrow{p} 1$, as for instance shown by Hillebrand (2005), but this issue is outside the scope of the present paper.

References

Baillie, R. T. and T. Bollerslev: 1989, 'The message in daily exchange rate: a conditional-variance tale'. *Journal of Business & Economic Statistics* 7, 60–68.

Baillie, R. T. and T. Bollerslev: 1990, 'Intra-Day and inter-market volatility in foreign exchange rates'. *Review of Economic Studies* 58, 565–585.

Baillie, R, T. and H. Chung: 2001, 'Estimation of GARCH models from the autocorrelations of the squares of a process'. *Journal of Time Series Analysis* 22, 631–650.

Bollerslev, T.: 1986, 'Modelling the persistence of conditional variances. With comments and a reply by the authors'. *Econometric Reviews* 5, 1–87.

Bollerslev, T. and H. O. Mikkelsen.: 1996, 'Modelling and pricing long memory in stock market volatility'. *Journal of Econometrics* 73, 151–184,.

Breit, T., N. Crato, and Lima: 1998, 'The detection and estimation of long memory in stochastic volatility'. *Journal of Econometrics* 83, 325–348.

Diebold, F.: 1986, 'Modelling the persistence of the conditional variances: a comment'. *Econometric Reviews* 5, 51–56.

Diebold, F. and A. Inoue: 2001, 'Long memory and regime switching'. *Journal of Econometrics* 105, 139–159.

Ding, Z., C. Granger, and R. Engle: 1993, 'A long memory property of stock market returns'. *Journal of Empirical Finance* 1, 83–106.

Dittmann, I. and C. Granger: 2002, 'Properties of nonlinear tranformations of fractionally integrated processes'. *Journal of Econometrics* 110, 113–133.

Dueker, M.: 1997, 'Markov switching in GARCH processes and mean-reverting stock-market volatility'. *Journal of Business & Economic Statistics* 15, 26–34.

Engle, R. and A. Smith: 1999, 'Stochastic permanent breaks'. *Review of Economics and Statistics* 81, 553–574.

Francq, C., J. Zakoian, and M. Roussignol: 2001, 'Conditional heteroscedasticity driven by hidden Markov chains'. *Journal of Time Series Analysis* 22, 197–220.

French, K. R., G. W. Schwert and Stambaugh, R. F.: 1987, 'Expected stock returns and volatility'. *Journal of Financial Economics* 19, 3–29

Haas, M., S. Mittnik, and M. S. Paollela: 2004, 'A new approach to Markov-switching GARCH models'. *Journal of Financial Econometrics* 2, 493–530.

Hamilton, J. D. and R. Susmel: 1994, 'Autoregressive conditional heteroscedasticity and changes in regime'. *Journal of Econometrics* 64, 307–333.

Hassler, U.: 1997, 'Sample autocorrelations of nonstationary fractionally integrated series'. *Statistical Papers* 38, 43–62.

Hillebrand, E.: 2005, 'Neglecting parameter changes in GARCH models'. *Journal of Econometrics* 129, 121–138.

Klaassen, F.: 2002, 'Improving GARCH volatility forecasts with regime-switching GARCH'. *Empirical Economics* 27, 363–394.

Krämer, W.: 1985, *Trend in Ökonometrischen Modellen.* Königstein, (Athenaeum/Hain).

Lamoreux, C., W. Lastrapes, and W. Lamoureux: 1990, 'Persistence in variance, structural change, and the GARCH Model'. *Journal of Business & Economic Statistics* 8, 225–234.

Mikosch, T. and C. Starica: 2004, 'Nonstationarities in financial time series, long range dependence, and IGARCH effects'. *Review of Economics and Statistics* 86, 378–390.

Newey, W. K. and K. D. West: 1987, 'A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix'. *Econometrica* 55, 703–708.

Wong, C. and W. Li: 2001, 'On a mixture autoregressive heteroscedastic model'. *Journal of the American Statistical Association* 96, 982–995.

Figure 1 Estimated persistence as a function of sample size

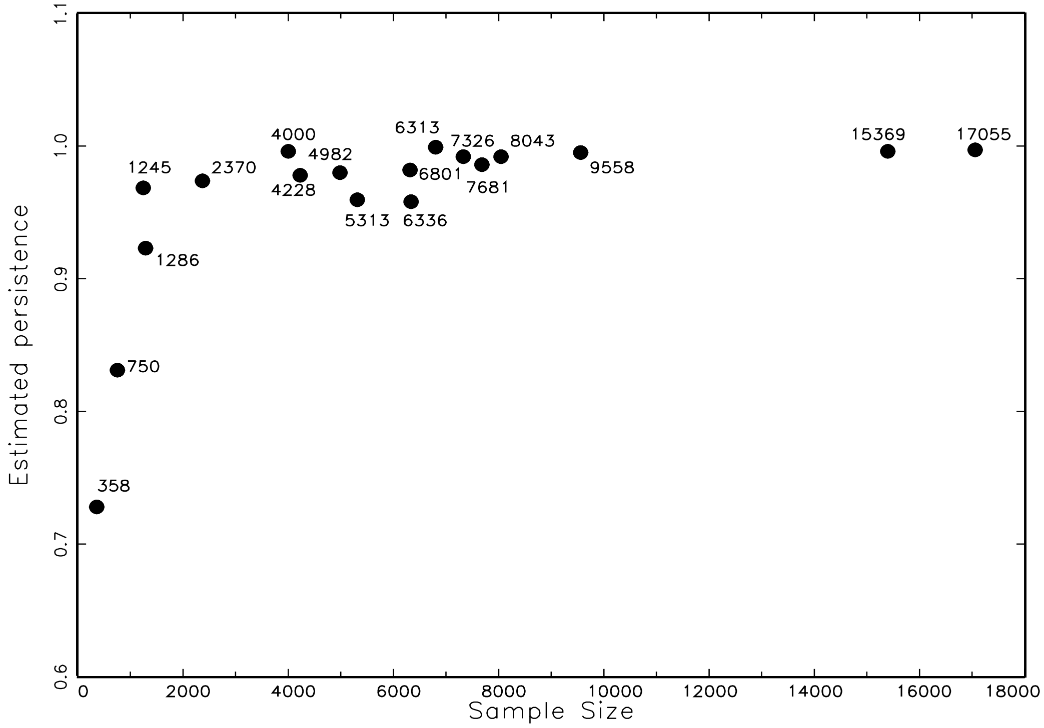
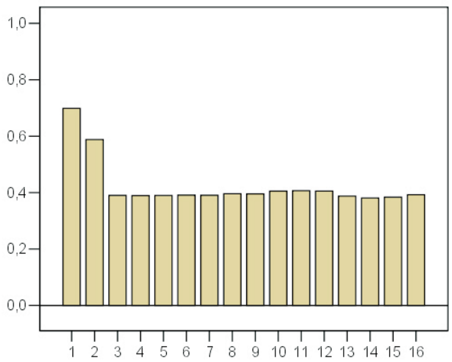
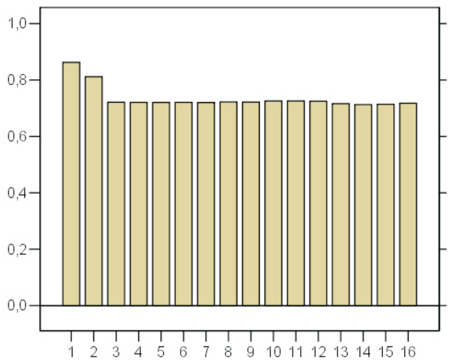


Figure 2 Empirical autocorrelations with a shift in expectations

a) d=1



b) d=2



c) d=4

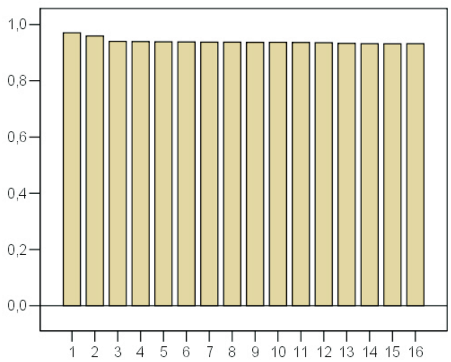


Figure 3 Estimated persistence as a function of the size of the break ($\alpha = 0.2$, $\beta = 0.4$)

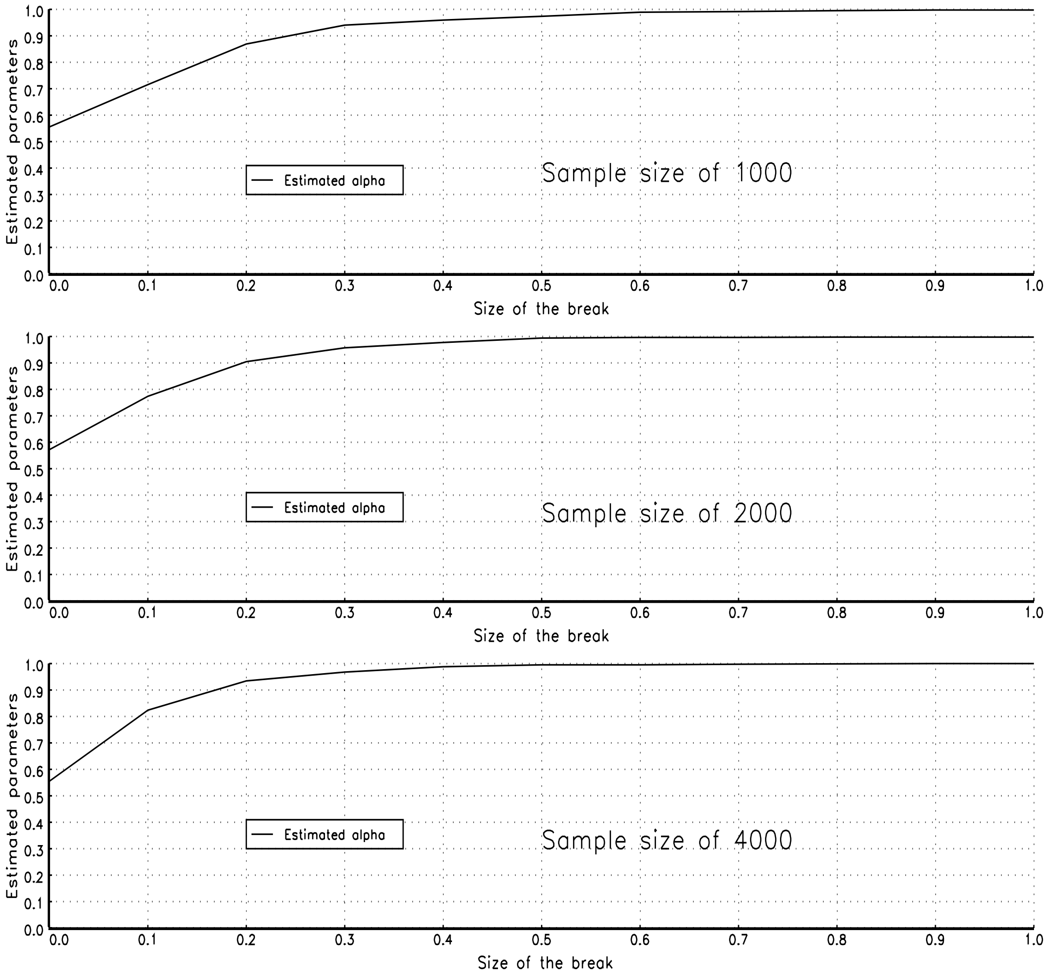


Figure 4 Estimated persistence as a function of the size of the break ($\alpha = \beta = 0.3$)

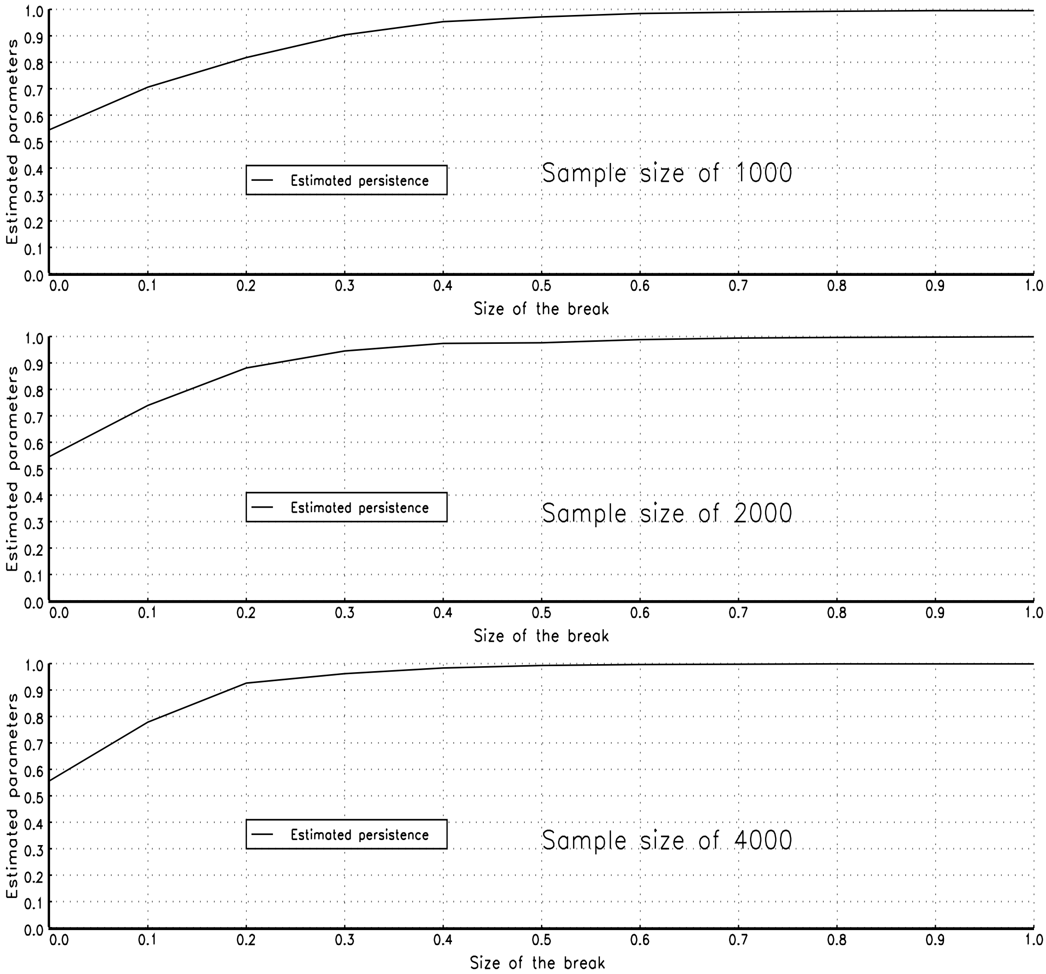


Figure 5 Estimated persistence as a function of the size of the break ($\alpha = 0.4,\ \beta = 0.2$)

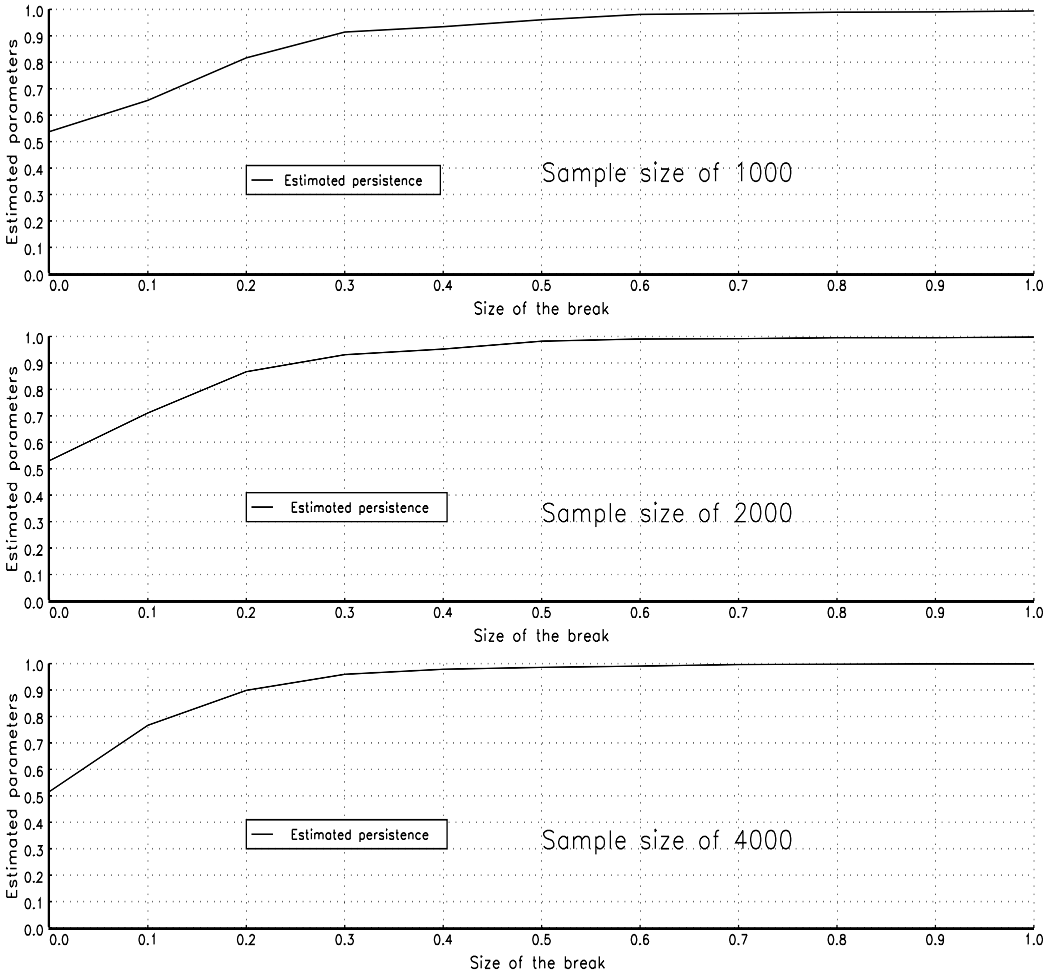
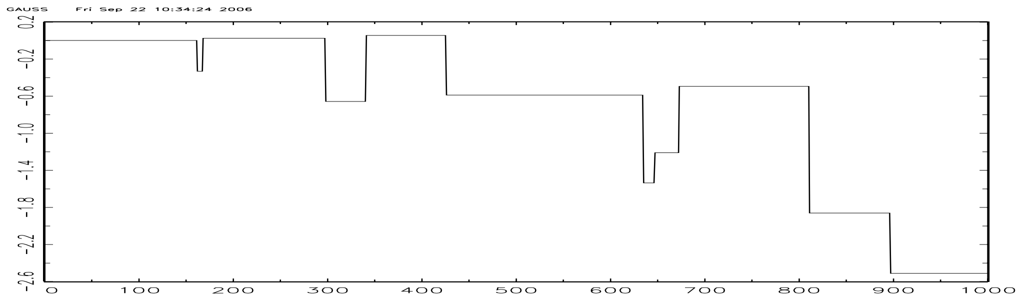
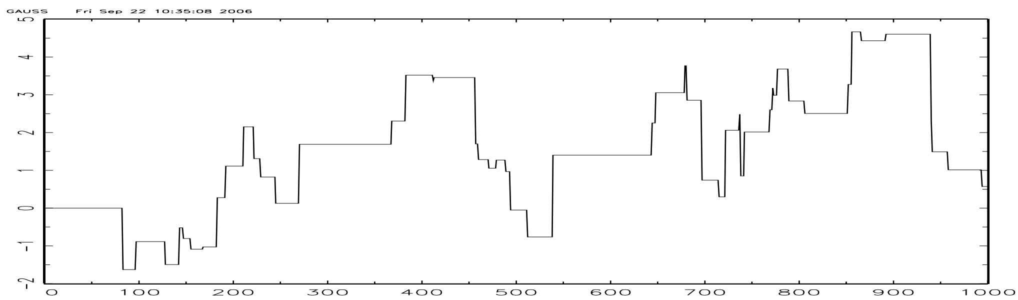


Figure 6 Stochastic mean according to equation (7), sample size = 1000

a) $p = 0.01$



b) $p = 0.05$



c) $p = 0.10$

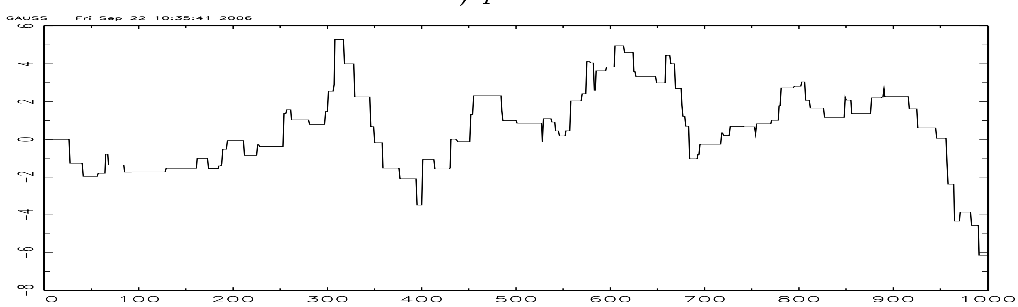
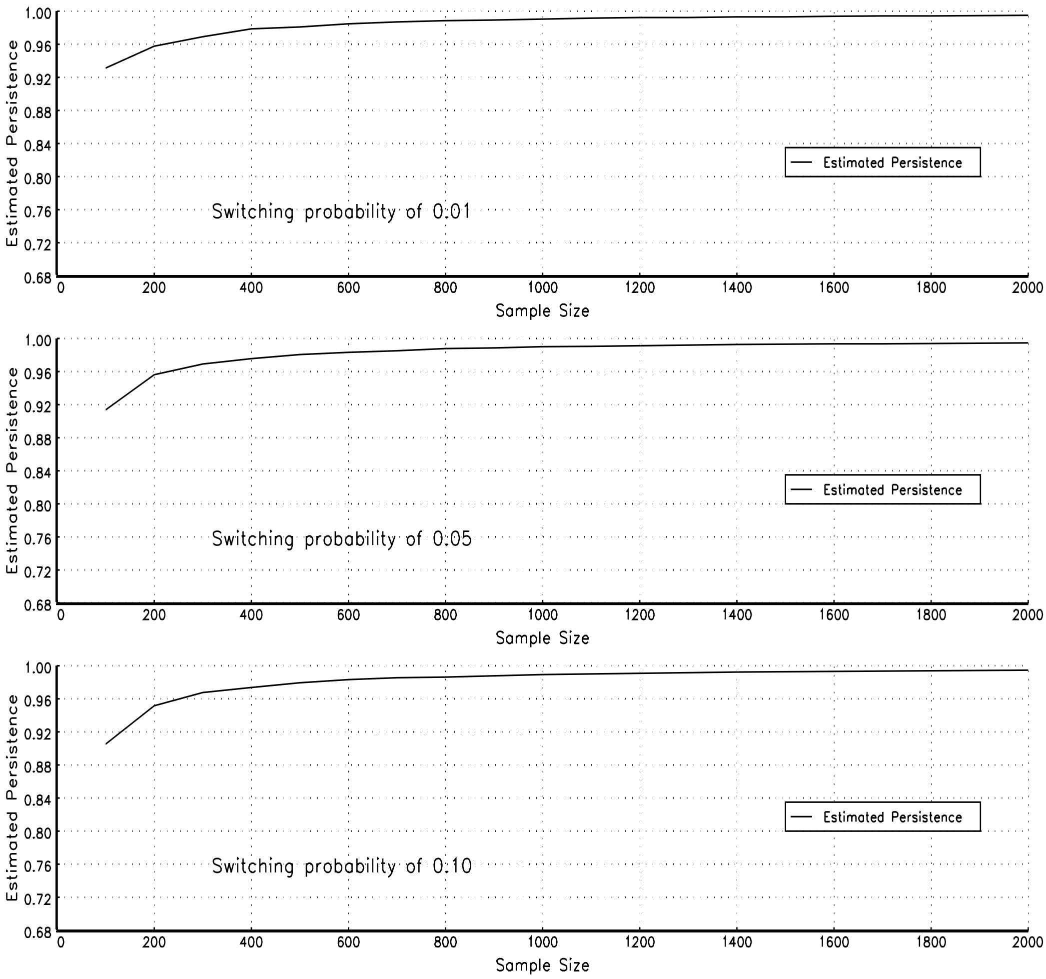


Figure 7 Estimated persistence as a function of the sample size with switching probabilities $p = 0.01$, 0.05 and 0.10 respectively



Appendix

Detailed description of empirical papers from figure 1

Authors	Data	Sample Size	$\hat{\delta}$
Lamoreux et al. (1990) *	20 stocks (80–84)	358	0.728
Mikosch and Starica (2004)	S+P 53-56	750	0.831
Baillie and Bollerslev (1989)	FX FF,US 01/03/80-28/01/85	1245	0.943
Baillie and Bollerslev (1989)	FX IT,US 01/03/80-28/01/85	1245	0.961
Baillie and Bollerslev (1989)	FX JPY,US 01/03/80-28/01/85	1245	0.990
Baillie and Bollerslev (1989)	FX CHF,US 01/03/80-28/01/85	1245	0.980
Baillie and Bollerslev (1989)	FX BP,US 01/03/80-28/01/85	1245	0.971
Baillie and Bollerslev (1989)	FX DM,US 01/03/80-28/01/85	1245	0.966
Francq et al. (2001)	CAC40 1/6/88-31/12/93	1286	0.923
Dueker (1997)	S+P Dec82-Dec91	2370	0.974
Hillebrand (2005)	D J 07/12/87-31/10/03	4000	0.996
Lamoreux et al. (1990) *	30 stocks 01/01/63-13/11/79	4228	0.978
Klaassen (2002) **	FX 03/01/78-23/07/97	4982	0.980
Haas et al. (2004)	FX SingD,USD 01/81-06/03	5313	0.933
Haas et al. (2004)	FX SingD,USD 01/81-06/03	5313	0.986
Haas et al. (2004)	FX BP,USD 01/81-06/03	6313	0.974
Haas et al. (2004)	FX BP,USD 01/81-06/03	6313	0.990
Haas et al. (2004)	FX JPY/USD 01/81-06/03	6336	0.958
Haas et al. (2004)	FX JPY/USD 01/81-06/03	6336	0.965
Breit et al. (1998)	VW Ret. Jul62-Jul69	6801	0.999
French and Schwert (1987)	S+P 01/28-12/52	7326	0.992
Haas et al. (2004)	Nasdaq Ret 02/71-06/01	7681	0.986

continued on next page

Authors	Data	Sample Size	$\hat{\delta}$
French and Schwert (1987)	S+P 01/53-12/84	8043	0.992
Bollerslev and Mikkelsen (1996)	S+P 500 02/01/53-31/12/90	9558	0.995
French and Schwert (1987)	S+P 01/28-12/84	15369	0.996
Ding et al. (1993)	S+P 500 03/01/28-30/08/91	17055	0.997

* average over 3 FX-rates

** average over 20 companies