

On the origins of high persistence in GARCH-models ¹

by

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

version June 2009

Summary

The estimated persistence in various types of GARCH - models is known to be too large when the parameters of the model undergo structural changes somewhere in the sample. The present paper adds further insights into this phenomenon for the Baillie and Chung (2001) minimum distance estimates of the model parameters. While previous research has focused on the effects of changes in the GARCH- parameters, we investigate here the consequences of a changing mean.

Keywords: minimum distance estimates, structural change, long memory, GARCH.

¹Research supported by Ruhr Graduate School of Economics, Dortmund-Bochum-Essen, and by Deutsche Forschungsgemeinschaft. We are grateful to Richard Baillie and Huimin Chung for providing the Gauss codes for the Minimum Distance estimator.

1 Introduction and summary

Among the large family of ARCH- and GARCH-models that have been proposed since the seminal paper of Engle (1982), the GARCH(1,1) - specification,

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

where the η_t are *iid*(0, 1) and independent of past ϵ 's and σ 's, is still by far the most popular. Typical applications include stock returns or inflation rates. However, it is often found in many applications that the estimate of the "persistence parameter" $\delta := \alpha + \beta$ appears as much too large, and that this upward bias increases as sample size increases.

It has long been known (see e.g. Diebold 1986) that this upward bias of the estimated persistence parameter might well be an artifact of structural change, either in μ or in the GARCH - parameters α , β and ω . Mikosch and Starica (2004) show that the Whittle-estimator of $\delta = \alpha + \beta$ must tend to 1 for any given sample size when the structural change increases, and Hillebrand (2005) proves the same for the ML- estimates for the case of given structural changes and increasing samples. The present paper considers the Baillie and Chung (2001) Minimum Distance estimator, extending Krämer and Tameze (2007). While Krämer and Tameze (2007) were mainly concerned with increasing structural changes for a given sample, in the vein of Mikosch and Starica (2004), we focus here on the empirically more relevant case of given changes and increasing samples, and show that the sum of the estimated α and β can likewise be made arbitrarily close to 1 if there are certain types of structural change in the x_t -process, in particular, structural changes in the expectation μ .

In the context of returns of risky assets, μ is the expected return, and there is no reason to believe that this remains constant over long stretches of time. The present analysis can therefore be viewed as an investigation of the consequences when such changes in expected returns are not properly accounted for.

2 Structural changes in the mean and sample autocorrelations

We first explore the relationship between structural change in the mean of model (1) and the empirical autocorrelations of the ϵ_t^2 (which are identical to the estimated autocorrelations of the x_t^2). These empirical autocorrelations are important because they provide the major input for the Minimum Distance estimator of α and β which we consider here.

Let therefore in general

$$z_t = \gamma_t + \eta_t \quad (t = 1, \dots, T)$$

be a sequence of random variables where the sequences $\{\gamma_t\}$ and $\{\eta_t\}$ are independent, with zero mean and weakly stationary η_t . If $\gamma_t = \gamma$ is fixed and nonstochastic, the h -th order autocorrelations ρ_h of $\{z_t\}$ and $\{\eta_t\}$ coincide and are consistently estimated by the empirical h 'th order autocorrelations of the z_t - sequence:

$$\hat{\rho}_h = \frac{\sum_{t=1}^{T-h} (z_t - \bar{z})(z_{t+h} - \bar{z})}{\sum_{t=1}^T (z_t - \bar{z})^2}. \quad (2)$$

These estimates can be affected by structural change in γ in various ways. One possibility, investigated in detail by Diebold and Inoue (2001), is real or spurious long memory in $\{\gamma_t\}$ (and therefore also in $\{z_t\}$), as defined by

$$\text{Var}\left(\sum_{t=1}^T z_t\right) = O(T^{2d+1}), \quad 0 < d \leq 1. \quad (3)$$

It easy to show (see e.g. Hassler 1997) that we must then have

$$p \lim_{T \rightarrow \infty} \hat{\rho}_h = 1 \quad (4)$$

for all h whenever the long-memory parameter d is larger than 1/2. Simply rewrite (2) as

$$\hat{\rho}_h = 1 - \frac{\sum_{t=T-h+1}^T (z_t - \bar{z})^2}{\sum_{t=1}^T (z_t - \bar{z})^2} + \frac{\sum_{t=1}^{T-h} (z_t - \bar{z})(z_{t+h} - z_t)}{\sum_{t=1}^T (z_t - \bar{z})^2} \quad (5)$$

and note that the numerators in the last two expressions are of smaller order in probability than the denominator:

$$\sum_{t=1}^T (z_t - \bar{z})^2 T \xrightarrow{P} \infty, \quad (6)$$

whereas both numerators are $Op(T)$.

Diebold and Inoue (2001) show that behavior of type (3) occurs for instance whenever γ_t is stochastic and independent of η_t and displays structural breaks of the form

$$\begin{aligned} \gamma_t &= \gamma_{t-1} + \nu_t \\ \nu_t &= \begin{cases} 0 & \text{with probability } 1-p \\ \omega_t & \text{with probability } p, \end{cases} \end{aligned} \quad (7)$$

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

$$\sum_{t=1}^T \gamma_t = T v_1 + (T-1)v_2 + \dots + v_T, \quad (8)$$

we have

$$Var\left(\sum_{t=1}^T \gamma_t\right) = p \sigma^2 \sum_{t=1}^T t^2 = p \sigma^2 \frac{T(T+1)(2T+1)}{6}, \quad (9)$$

so we can have (3) for any d , $0 < d < 1$, by letting p vary with sample size according to

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

A similar way of introducing long memory in $\{z_t\}$ is letting γ takes values γ_1 and $\gamma_2 (\gamma_1 \neq \gamma_2)$ on consecutive intervals of some renewal process (Leipus and Surgailis 2003). If the renewal distribution has heavy tails, there will be long memory in $\{z_t\}$.

Below we focus on a second type of structural change which likewise induces spurious long memory in $\{z_t\}$, which is given by r nonstochastic shifts in mean at

$$Tq_j (q_j = p_0 + p_1 + \dots + p_j, \quad p_i > 0, \quad \sum p_i = 1), \quad (11)$$

so $\gamma_t = \gamma_i$ whenever $Tq_i < t \leq T(q_i + 1)$. There appears to emerge a consensus in empirical finance that this type of structural change is the predominant one in the context of squared stock returns (Starica and Granger 2005). With such nonstochastic changes, it can easily be shown (see e.g. Mikosch and Starica 2004) that, for fixed h and $T \rightarrow \infty$, we have

$$\hat{\rho}_h \xrightarrow{p} \frac{\rho_h \text{var}(\eta_t) + \sum_{i,j=0}^r (\gamma_i - \gamma_j)^2}{\text{var}(\eta_t) + \sum_{i,j=0}^r (\gamma_i - \gamma_j)^2}, \quad (12)$$

whereas for given T and increasing structural changes, i.e. $\sum_{i,j=0}^r (\gamma_i - \gamma_j)^2 \rightarrow \infty$, we have

$$\hat{\rho}_h \xrightarrow{p} 1. \quad (13)$$

The latter relationship is the crucial element in the proof in Mikosch and Starica (2004) and Krämer and Tameze (2007) that the Whittle and Minimum Distance estimators of $\alpha + \beta$ likewise tend to 1 for given T as structural changes are increasing.

One might question, however, whether such extreme structural changes are really relevant in applications. As will be shown below, it suffices for the Minimum Distance estimator to tend to one that the limit in (12) remains bounded away from zero as $h \rightarrow \infty$:

$$\lim_{h \rightarrow \infty} \frac{\rho_h \text{var}(\eta_t) + \sum_{i,j=0}^r (\gamma_i - \gamma_j)^2}{\text{var}(\eta_t) + \sum_{i,j=0}^r (\gamma_i - \gamma_j)^2} = \frac{\sum_{i,j=0}^r (\gamma_i - \gamma_j)^2}{\text{var}(\eta_t) + \sum_{i,j=0}^r (\gamma_i - \gamma_j)^2} > 0. \quad (14)$$

3 The Minimum Distance estimator with structural change

Next we set $z_t = x_t^2$ and consider the Baillie and Chung (2001) Minimum Distance estimator of the GARCH - parameters α and β in the model (1),

given that empirical autocorrelations of the x_t^2 behave as explained in section 2. This estimator exploits the fact that the ϵ_t^2 can be written as an ARMA(1,1) - process

$$\epsilon_t^2 = \omega + (\alpha + \beta)\epsilon_{t-1}^2 + u_t - \beta u_{t-1}, \quad (15)$$

where

$$u_t := \epsilon_t^2 - E(\epsilon_t^2 | \epsilon_{t-1}^2, \epsilon_{t-2}^2, \dots) = \epsilon_t^2 - \sigma_t^2 \quad (16)$$

is white noise and uncorrelated with past ϵ_t^2 's, and that theoretical autocorrelations of ϵ_t^2 are therefore known functions of α and β :

$$\begin{aligned} \rho_1 &= \alpha + \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2} \\ \rho_2 &= \left(\alpha + \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2}\right)(\alpha + \beta) \\ &\vdots \\ \rho_h &= \left(\alpha + \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2}\right)(\alpha + \beta)^{h-1} \quad (h > 1). \end{aligned} \quad (17)$$

The Minimum Distance Estimators $\hat{\alpha}$ and $\hat{\beta}$ for α and β are then obtained as

$$\arg \min_{\alpha, \beta} [\hat{\rho} - \rho(\alpha, \beta)]' W [\hat{\rho} - \rho(\alpha, \beta)], \quad (18)$$

where W is some suitable positive definite weighting matrix, $\hat{\rho} = (\hat{\rho}_1, \dots, \hat{\rho}_h)'$ is the vector of estimated autocorrelations and where $\rho(\alpha, \beta) = (\rho_1, \dots, \rho_h)'$ is the vector-valued function of α and β defined in (17).

Krämer and Tameze (2007) consider the case where the $\hat{\rho}_i$ tend to one, i.e where

$$\hat{\rho} = (\hat{\rho}_1, \dots, \hat{\rho}_h)' \xrightarrow{P} e := (1, \dots, 1)', \quad (19)$$

and show that this implies that $\hat{\alpha} + \hat{\beta} \xrightarrow{P} 1$. From

$$\begin{aligned} plim(\hat{\alpha}, \hat{\beta}) &= \arg \min_{\alpha, \beta} [plim \hat{\rho} - \rho(\alpha, \beta)]' W [plim \hat{\rho} - \rho(\alpha, \beta)] \\ &\subseteq \arg \min_{\alpha, \beta} [e - \rho(\alpha, \beta)]' W [e - \rho(\alpha, \beta)], \end{aligned} \quad (20)$$

one sees that the latter set of minimizing values of α and β is in view of (17) determined by

$$\alpha + \beta = 1 \quad \text{and} \quad (21)$$

$$\alpha + \frac{\alpha^2\beta}{1 - 2\alpha\beta - \beta^2} = 1, \quad (22)$$

for which $\rho(\alpha, \beta) = e$, so

$$[e - \rho(\alpha, \beta)]'W[e - \rho(\alpha, \beta)] = 0, \quad (23)$$

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

The empirical usefulness of this results is rather limited, however. As $\hat{\rho} \rightarrow e$ requires either nonstationary long memory ($d > 1/2$) as sample size increases or unlimited structural changes when sample size is fixed, one would rarely expect this to happen in e.g. financial applications. Much more relevant are small but persistence structural changes of the type (3), where, in view of (14),

$$\hat{\rho}_h \rightarrow q \quad (0 < q < 1) \quad (24)$$

as sample size increases. Next we show that, also in this case, the Minimum Distance estimator of $\alpha + \beta$ must by logical necessity tend to 1.

THEOREM:

Whenever the number h of empirical autocorrelations of the x_t^2 which is used for the Minimum Distance estimators $\hat{\alpha}$ and $\hat{\beta}$ of α and β tends to infinity as sample size increases, the relationship (24) implies $p \lim(\hat{\alpha} + \hat{\beta}) = 1$.

PROOF:

The h - dimensional vector $\rho(\hat{\alpha}, \hat{\beta})$, with typical element

$$\rho_i = \left(\hat{\alpha} + \frac{\hat{\alpha}^2\hat{\beta}}{1 - 2\hat{\alpha}\hat{\beta} - \hat{\beta}^2}\right)(\hat{\alpha} + \hat{\beta})^i \quad (i = 1, \dots, h), \quad (25)$$

which solves the minimization problem (18) and therefore yields the Minimum Distance estimator $\hat{\alpha}$ and $\hat{\beta}$, is componentwise geometrically decreasing in i . Therefore, its final element must eventually obey the restriction

$$\left(\hat{\alpha} + \frac{\hat{\alpha}^2 \hat{\beta}}{1 - 2\hat{\alpha}\hat{\beta} - \hat{\beta}^2}\right)(\hat{\alpha} + \hat{\beta})^h > q - \varepsilon$$

for any $\varepsilon > 0$. Taking the h -th root on both sides of the inequality yields

$$\left(\hat{\alpha} + \frac{\hat{\alpha}^2 \hat{\beta}}{1 - 2\hat{\alpha}\hat{\beta} - \hat{\beta}^2}\right)^{1/h}(\hat{\alpha} + \hat{\beta}) > (q - \varepsilon)^{1/h},$$

where the right hand side tends to 1 as $h \rightarrow \infty$. Therefore, the left hand side must tend to one as well, which in turn implies $\hat{\alpha} + \hat{\beta} \xrightarrow{p} 1$.

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

$$\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha}\hat{\varepsilon}_{t-1}^2 + \hat{\beta}\hat{\sigma}_{t-1}^2 \quad (t = 1, \dots, T), \quad (26)$$

where the $\hat{\sigma}_t^2$ and $\hat{\varepsilon}_t$ are fitted values obtained from the ML-estimator for ω , α and β and some starting values ε_0^2 and σ_0^2 . If there are in addition only a fixed number of regimes, with regime-specific expectations $E(\sigma^2)_{(i)} = E(\varepsilon_t^2)_{(i)} =: E_i$ and with regime-specific sample sizes increasing, one obtains under certain conditions on the estimators that

$$\overline{\hat{\sigma}^2}_{(i)} \xrightarrow{p} E_{(i)}, \overline{\hat{\varepsilon}^2}_{(i)} \xrightarrow{p} E_{(i)}, \quad (27)$$

so

$$E_{(i)} - E \approx (\hat{\alpha} + \hat{\beta})(E_{(i)} - E) \quad (28)$$

where E is the sample mean of the $\hat{\sigma}^2$. Therefore $\hat{\alpha} + \hat{\beta}$ must likewise tend to 1. This argument however depends crucially on the validity of the limiting relationship in (27) and is different from the one advanced in the present paper.

4 Some finite sample simulations

This section reports on various Monte Carlo simulations to check the finite sample relevance of the above results. In a first series of experiments, we keep the number of changes fixed at times $[Tq_1][Tq_2], \dots, [Tq_k]$ where $0 < d_1 < d_2 < \dots < d_r < 1$, along the lines of Starica and Granger (2005), and Hillebrand (2005).

Figure 1 reports the first 35 empirical autocorrelations of a GARCH(1,1)-process with $\alpha = 0.2$, $\beta = 0.4$, $\omega = 0.001$, $\eta_t \sim n.i.d(0, 1)$ where $r = 1$, $d_1 = 1/2$, and where a shift in μ of size 0.8 occurs in the middle of the sample. The figures are averages over 1000 replications. The figure also indicates the limit q from equation (24) and shows that the limit is approached quick rapidly as h increases, also for modest values of T .

Table 1 shows the resulting estimates of $\hat{\delta} = \hat{\alpha} + \hat{\beta}$, also for a wider range of sample sizes and structural breaks. It is seen that the estimated persistence likewise tends to 1 quite rapidly as the sample size increases, at least if the structural change is large enough, and that $\hat{\delta}$ is biased downwards in small samples if there is no structural change. This downward bias vanishes very slowly as sample size increases. Similar results were also obtained for other values of r, d_1, d_2, \dots, d_r and α and β and can be obtained from the authors upon request.

In a second series of experiments, we let μ change according to the Diebold and Inoue (2001)-scheme from equation (7). Figure 2 shows the resulting first 35 empirical autocorrelations of the x_t^2 for the case where $\omega_t \sim n.i.d(0, 1)$ and the switching probability is $p = 0.05$. It is seen that sample autocorrelations tend to a constant as sample size increases.

Table 2 gives the persistence in a GARCH(1,1)-model derived from these empirical autocorrelations for a wider range of switching probabilities where $\nu_t \sim n.i.d(0, 1)$ and sample sizes n . Again, it is seen that $\hat{\delta}$ approaches 1 quite rapidly, and similar results were obtained for different parameters of the GARCH-model as well.

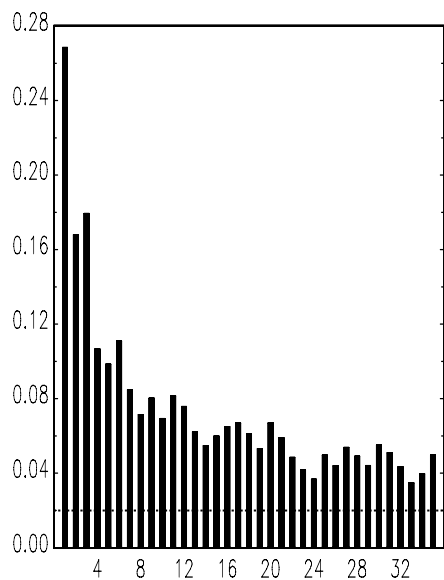
5 Conclusion

The present paper confirms the conventional wisdom that overly large estimated persistence in GARCH-models is not necessarily due to a large real persistence. Rather, it might as well be structural changes in the model parameters. Extending previous work which was mostly concerned with changes in the GARCH -parameters, we show here how changes in this expectation might likewise bias the estimated persistence towards unity.

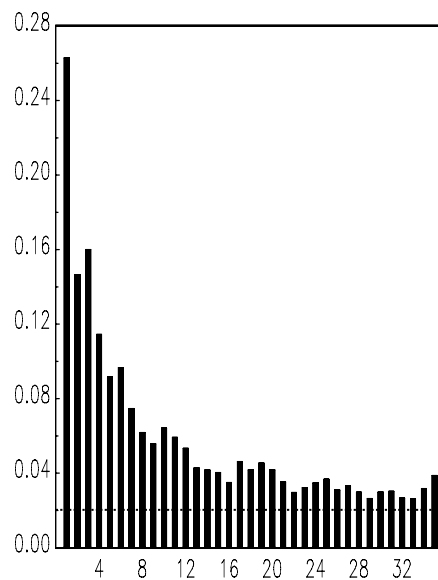
References

- Baillie, R. T. and H. Chung (2001): ‘Estimation of GARCH models from the autocorrelations of the squared of a process’. *Journal of Time Series Analysis* **22**(6), 631–650.
- Diebold, F. (1986): ‘Modeling 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.
- Engle R. (1982): ‘Autoregressive Conditional Heterokedasticity with estimates of the variance of the United Kingdom Inflation’. *Econometrica* (50), 987–1007.
- 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.
- Krämer W. and B. Tameze (2007): ‘Structural change and estimated persistence in the GARCH(1,1)-model’. *Economics Letters* **97**(1), 17–23.
- Leipus, R. and D. Surgailis (2003): ‘Random coefficient autoregression, regime switching and long memory’. *Applied Probability* (35), 737–754.
- Mikosch, T. and C. Starica (2004): ‘Nonstationarities in Financial Times Series, the Long Range Dependence, and the IGARCH Effects’. *The Review of Economics and Statistics* (86), 378–390.
- Starica, C. and C. Granger (2005): ‘Nonstationarities in stock returns’. *The Review of Economics and Statistics* (87), 503–522.

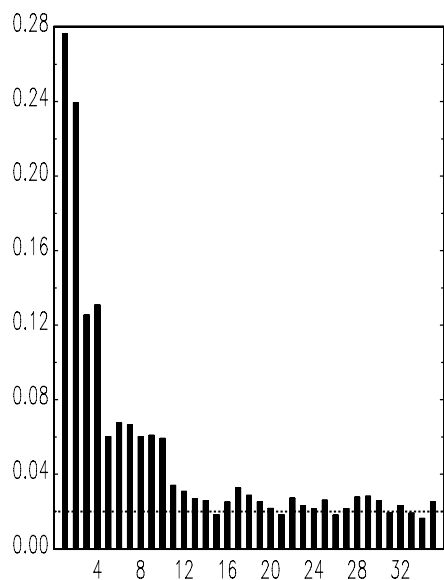
Figure 1: Sample autocorrelations in the context of a nonstochastic change in mean in the middle of the sample



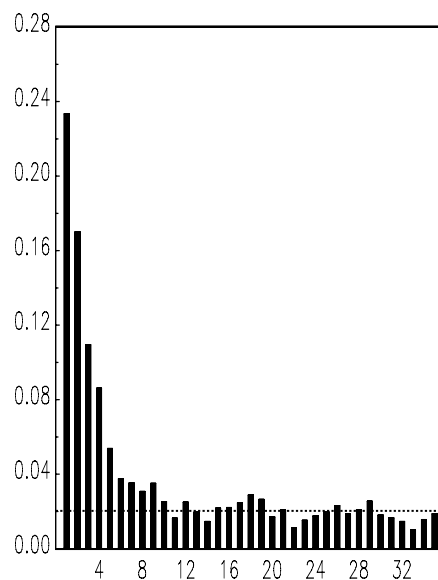
(a) $T=500$



(b) $T=1000$

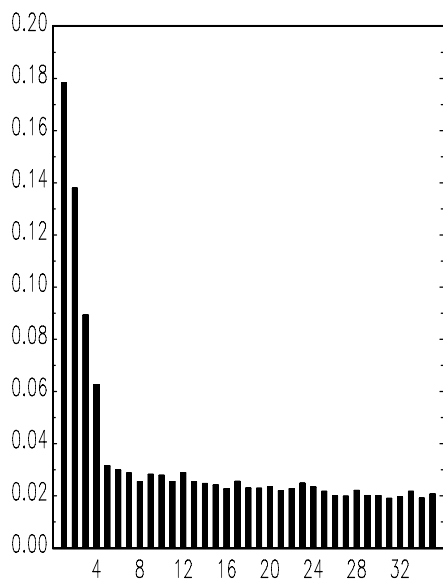


(c) $T=2000$

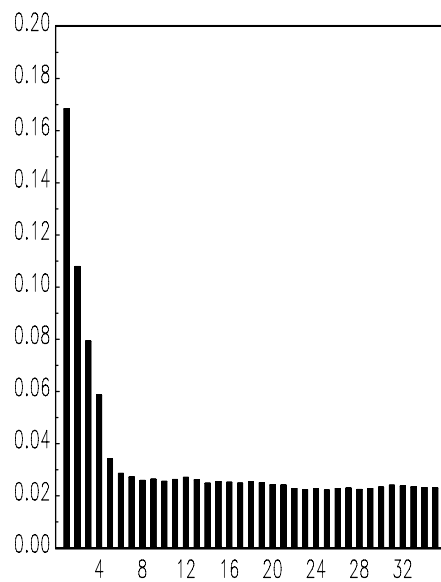


(d) $T=4000$

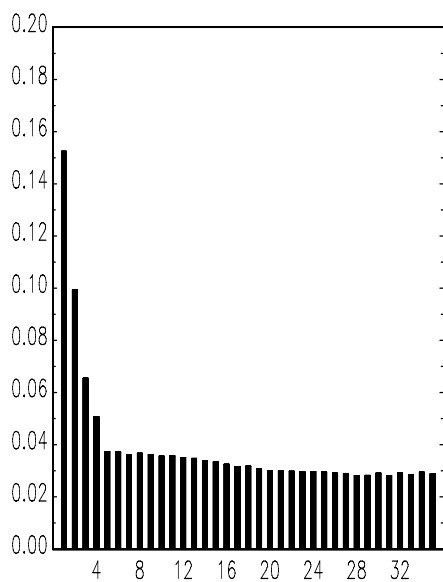
Figure 2: Sample autocorrelations in the context of a stochastic change and switching probability of $p = 0.05$



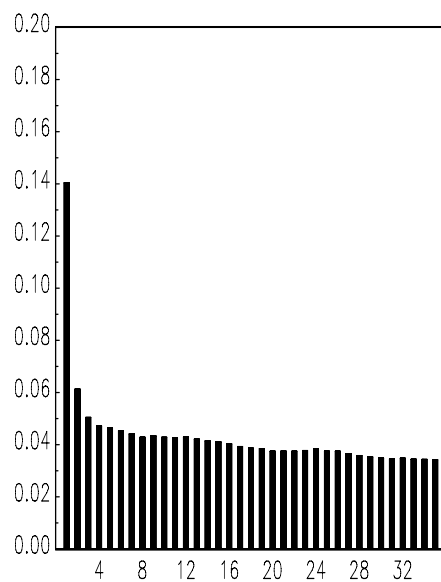
(e) $T=500$



(f) $T=1000$



(g) $T=2000$



(h) $T=4000$

Table 1: Expected values of $\hat{\delta} = \hat{\alpha} + \hat{\beta}$ for various changes in mean and sample size

T	$\Delta\mu$							
	0	0.2	0.4	0.6	0.8	1	1.2	1.4
a) $\alpha = 0.2, \beta = 0.4$								
500	0.5754	0.8514	0.9304	0.9681	0.9842	0.9907	0.9950	0.9953
1000	0.5825	0.8638	0.9507	0.9853	0.9929	0.9959	0.9981	0.9983
2000	0.5912	0.8903	0.9767	0.9943	0.9972	0.9984	0.9989	0.9992
4000	0.5976	0.9374	0.9878	0.9977	0.9992	0.9994	0.9995	0.9996
b) $\alpha = 0.4, \beta = 0.2$								
500	0.5634	0.7849	0.8949	0.9482	0.9792	0.9873	0.9917	0.9932
1000	0.5762	0.8009	0.9241	0.9711	0.9872	0.9930	0.9935	0.9978
2000	0.5892	0.8593	0.9577	0.9868	0.9929	0.9966	0.9969	0.9990
4000	0.5947	0.9037	0.9728	0.9939	0.9975	0.9991	0.9984	0.9994
c) $\alpha = 0.3, \beta = 0.3$								
500	0.5714	0.8115	0.9059	0.9629	0.9839	0.9882	0.9907	0.9932
1000	0.5796	0.8405	0.9420	0.9763	0.9895	0.9917	0.9960	0.9974
2000	0.5883	0.8871	0.9709	0.9887	0.9954	0.9969	0.9988	0.9991
4000	0.5964	0.9093	0.9840	0.9953	0.9986	0.9994	0.9995	0.9996

Table 2: Expected values of $\hat{\delta} = \hat{\alpha} + \hat{\beta}$ for various switching probabilities and sample size

T	<i>p</i>										
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
a) $\alpha = 0.2, \beta = 0.4$											
500	0.572	0.754	0.858	0.931	0.971	0.982	0.986	0.994	0.995	0.996	0.996
1000	0.584	0.784	0.884	0.954	0.968	0.989	0.995	0.996	0.997	0.997	0.998
2000	0.594	0.790	0.936	0.982	0.995	0.998	0.997	0.998	0.998	0.998	0.999
4000	0.597	0.857	0.962	0.984	0.997	0.998	0.998	0.999	0.999	0.999	0.999
a) $\alpha = 0.4, \beta = 0.2$											
500	0.563	0.700	0.795	0.906	0.946	0.967	0.973	0.992	0.993	0.992	0.993
1000	0.574	0.739	0.858	0.923	0.980	0.982	0.992	0.995	0.997	0.997	0.998
2000	0.589	0.755	0.919	0.953	0.965	0.990	0.993	0.998	0.997	0.998	0.999
4000	0.595	0.756	0.938	0.980	0.996	0.997	0.997	0.998	0.998	0.999	0.999
c) $\alpha = 0.3, \beta = 0.3$											
500	0.573	0.706	0.823	0.928	0.943	0.975	0.976	0.985	0.994	0.996	0.996
1000	0.581	0.726	0.922	0.961	0.964	0.983	0.995	0.997	0.997	0.998	0.998
2000	0.589	0.743	0.933	0.969	0.980	0.996	0.996	0.998	0.998	0.998	0.999
4000	0.596	0.829	0.953	0.984	0.995	0.997	0.998	0.998	0.999	0.999	0.999