Outlier detection in the GARCH(1,1) model

Philip Hans Franses*
Econometric Institute, Erasmus University Rotterdam

Dick van Dijk†
Econometric Institute, Erasmus University Rotterdam

July 5, 1999

Abstract

In this paper the issue of detecting and handling outliers in the GARCH(1,1) model is addressed. Simulation evidence shows that neglecting even a single outlier has a dramatic on parameter estimates. To detect and correct for outliers, we propose an adaptation of the iterative in Chen and Liu (1993, JASA). We generate the critical values for the relevant test statistic, and we evaluate our method in an extensive simulation study. An application to several weekly stock return series shows that correcting for a few outliers yields substantial improvements in out-of-sample forecasts.

Keywords: Autoregressive conditional heteroskedasticity, outliers, forecasting volatility.

*Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR, Rotterdam, The Netherlands, email: franses@few.eur.nl

†Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR, Rotterdam, The Netherlands, email: djvandijk@few.eur.nl (corresponding author)

1 Introduction

The Generalized Autoregressive Conditional Heteroskedasticity [GARCH] model often is applied to describe and forecast volatility clustering in financial returns data, see Bollerslev, Chou and Kroner (1992). Typically, the GARCH model which is considered in practice is of order (1,1) which, roughly speaking, corresponds with an ARMA(1,1) model for the conditional variance, see Bollerslev (1988). Volatility clusters usually emerge as sequences of outliers. Additional to these outliers, it may occur that financial returns series contain other atypical observations such as additive or innovation outliers, see Hotta and Tsay (1998) for a recent classification of outliers in GARCH models.. Economic arguments for the possible presence of atypical observations are given in, for example, Friedman and Laibson (1989).

In this paper, we address the issue of detecting and handling outliers in the GARCH(1,1) model. For this purpose, we adapt the outlier detection method proposed in Chen and Liu (1993). In a recent paper, Franses and Ghijsels (1999) apply this method for similar purposes, but these authors consider an inconvenient definition of oultiers in the context of a GARCH model. Here we consider a more naturally defined contamination model, based on one of the representations in Hotta and Tsay (1998). Sakata and White (1998) also focus on outliers in GARCH models using robust estimation methods, whereas we use the iterative scheme of Chen and Liu (1993), which amounts to a method which is rather easy to use.

The outline of this paper is as follows. In Section 2, we use Monte Carlo simulations to examine the impact of neglecting outliers on parameter estimates in the GARCH(1,1) model. We find that neglecting even a single outlier (of moderate magnitude) already has a dramatic impact on these parameter estimates. In Section 3, we put forward our adaptation of the Chen and Liu (1993) method for the case of outliers in the GARCH(1,1) model. We generate critical values of the relevant test statistic and we evaluate its empirical performance in an extensive simulation study. In Section 4, we apply our method to 6 weekly returns series for international stock indexes. We find that there are only a few outliers in each series, but that these have significant impact on the parameter estimates. Also, when we take care of these outliers, we obtain substantial improvements in out-of-sample forecasts of conditional volatility. In Section 5, we conclude with some remarks.

2 The effect of neglecting outliers

To gauge the importance of taking care of aberrant observations in a GARCH process, we consider the effect of neglecting a single outlier on the parameter estimates in a GARCH(1,1) model.

For this purpose, we consider the following model for an observed time series e_t, consisting of a GARCH(1,1) process ε_t which is contaminated with outliers, that is,

$$e_t = \varepsilon_t + \omega \cdot \text{sgn}(\varepsilon_t) I_t[t = \tau], \qquad t = 1, \ldots, n, \tag{1}$$

$$\varepsilon_t = z_t \sqrt{h_t}, \tag{2}$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{3}$$

where sgn is the sign function, $I_t[t = \tau] = 1$ if $t = \tau$ and zero otherwise, ω is a non-zero constant, z_t is an i.i.d. process with zero mean and unit variance, and n denotes the sample size. Hence, e_t is for example a financial returns series with an aberrant observation at time τ. As ε_t is assumed to be uncorrelated with its own past, this aberrant observation can be labeled both an additive and an innovation outlier. Below, we will return to this issue in more detail.

We perform the following experiment to examine the impact of neglecting the outlier at time τ on the maximum likelihood [ML] estimates of the parameters α_0, α_1 and β_1. Artificial series ε_t are generated according to (2)-(3), with $(\alpha_1, \beta_1) = (0.2, 0.5)$, (0.2,0.7), (0.1,0.6) and (0.1,0.8), which are typical values encountered in practice. The intercept α_0 is set equal to $\alpha_0 = 1 - \alpha_1 - \beta_1$, such that the unconditional variance of ε_t is equal to 1 for all combinations of α_1 and β_1. The innovations z_t are drawn from a standard normal distribution. We generate series ε_t, $t = -250, -249, \ldots, 0, 1, \ldots, n$ with $n = 250$ or 500. The necessary starting values ε_{-251}^2 and h_{-251} in (3) are set equal to their unconditional expectation, that is, equal to 1. The first 250 observations of each series are discarded to avoid that the results depend on these initial conditions. Next, the series e_t is obtained by adding a single outlier according to (1) at $\tau = \lfloor \pi n \rfloor$, where $\lfloor \cdot \rfloor$ denotes integer part. We set $\pi = 0.50$. Results for other values of π suggest the same qualitative conclusions. The magnitude of the outlier is varied among $\omega \in \{0, 3, 5, 7, 9\}$. Estimates of the parameters of a GARCH(1,1) model for the observed e_t series are obtained using maximum likelihood with analytic gradients, as advocated by Fiorentini, Calzolari and Panatoni (1996). The starting values for the parameters are determined from the first 5 autocorrelations of e_t^2.

Figures 1-6 show (kernel estimates of) the empirical distributions of the different parameter estimates, based on 1000 replications.

- insert Figures 1-6 about here -

Figures 1-3 are concerned with $n = 250$, and Figures 4-6 with $n = 500$. Figures 1 and 4 display the outcomes for α_0, Figures 2 and 5 for α_1, and Figures 3 and 6 for β_1.

The results for the intercept parameter α_0 indicate that the parameter estimates get less precise when the magnitude of the neglected outlier increases. Apparently, this holds for both sample sizes and for all four combinations of α_1 and β_1 considered in the simulation experiment.

Neglecting a single outlier has a dramatic impact on the estimate of α_1. This estimate clearly is biased towards zero, especially for large outliers. Again this holds for all four sets of parameter and for the two sample sizes.

Finally, the empirical distributions depicted in Figures 3 and 6 indicate that for a small sample, the estimate of β_1 is biased towards unity. For the larger sample size, we do not observe such a clear-cut tendency.

In sum, neglecting a single outlier has a substantial impact on parameter estimates, even in samples as large as 500 observations. These simulation results clearly suggest the need for a method that can detect and handle aberrant observations in a GARCH(1,1) model.

3 Outlier detection

In this section we outline the outlier detection method. We generate critical values for the test statistic and we examine the empirical performance of our method.

3.1 Method

Consider the following model for an observed time series e_t, consisting of a GARCH(1,1) process ε_t contaminated with an outlier

$$e_t = \varepsilon_t + \omega I_t[t = \tau], \tag{4}$$

$$\varepsilon_t = z_t \sqrt{h_t}, \tag{5}$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{6}$$

where the variables are as defined before. The parameters in the model for the conditional variance h_t of ε_t are assumed to satisfy the restrictions $\alpha_0 > 0$, $\alpha_1 > 0$, $\beta_1 > 0$ and $\alpha_1 + \beta_1 < 1$, such that the process is second-order stationary, see Bollerslev (1986). The GARCH(1,1) model (6) can be rewritten as an ARMA(1,1) model for ε_t^2,

$$\varepsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1}, \tag{7}$$

where $\nu_t = \varepsilon_t^2 - h_t$. Franses and Ghijsels (1999) take (7) as the starting point for their detection method. This means that implicitly the outlier is defined in terms of the squares of the observed time series, as

$$e_t^2 = \varepsilon_t^2 + \omega^* I_t[t = \tau].$$

The definition of an outlier as in (4) is however a more natural one. Furthermore, define the lag polynomial $\pi(L) = 1 - \pi_1 L - \pi_2 L^2 - \ldots$ as

$$\begin{aligned} \pi(L) &= (1 - \beta_1 L)^{-1}(1 - (\alpha_1 + \beta_1)L) \\ &= (1 + \beta_1 L + \beta_1^2 L^2 + \beta_1^3 L^3 + \ldots)(1 - (\alpha_1 + \beta_1)L) \\ &= 1 - \alpha_1 L - \alpha_1 \beta_1 L^2 - \alpha_1 \beta_1^2 L^3 - \ldots . \end{aligned} \tag{8}$$

Hence, $\pi_j = \alpha_1 \beta_1^{j-1}$ for $j = 1, 2, \ldots$. This allows (7) to be written as

$$\nu_t = \pi(L)\varepsilon_t^2 - \alpha_0/(1 - \beta_1).$$

This expression is used extensively below.

To examine the effects of outliers in this context, consider the case where a single outlier occurs at $t = \tau$, that is, the observed time series $e_t = \varepsilon_t$ for all $t = 1, \ldots, n$ with $t \neq \tau$, while $e_\tau = \varepsilon_\tau + \omega$. The conditional variance h_t is unobserved, but it can be estimated as

$$h_t^e = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 h_{t-1}^e. \tag{9}$$

For simplicity, we assume that the parameters α_0, α_1 and β_1 in the GARCH(1,1) model (6) are known. In practice these parameters are of course unknown and have to be estimated from the data. Under the null hypothesis of no outlier contamination, (quasi) ML estimates of the GARCH parameters are consistent and asymptotically normal, see Lee and Hansen (1994) and Lumsdaine (1996).

Under the assumptions made above, we have $h_t^e = h_t$ for $t \leq \tau$, and

$$\begin{aligned} h_{\tau+1}^e &= \alpha_0 + \alpha_1 e_\tau^2 + \beta_1 h_\tau^e \\ &= \alpha_0 + \alpha_1(\varepsilon_\tau + \omega)^2 + \beta_1 h_\tau \\ &= h_{\tau+1} + \alpha_1(\omega^2 + 2\omega\varepsilon_\tau), \\ h_{\tau+2}^e &= \alpha_0 + \alpha_1 e_{\tau+1}^2 + \beta_1 h_{\tau+1}^e \\ &= \alpha_0 + \alpha_1 \varepsilon_{\tau+1}^2 + \beta_1(h_{\tau+1} + \alpha_1(\omega^2 + 2\omega\varepsilon_\tau)) \\ &= h_{\tau+2} + \beta_1\alpha_1(\omega^2 + 2\omega\varepsilon_\tau), \end{aligned}$$

and in general

$$h_{\tau+j}^e = h_{\tau+j} + \beta_1^{j-1}\alpha_1(\omega^2 + 2\omega\varepsilon_\tau) \qquad \text{for } j = 1, 2, \ldots. \tag{10}$$

When computing the conditional variance according to (9) from the observed time series $e_1, \ldots, e_n$, certain assumptions have to be made concerning the necessary starting values in the recursion. Usual practice is to set both e_0 and h_0^e equal to the sample mean of the squared series, $\frac{1}{n}\sum_{t=1}^n e_t^2$. In case the time series is observed with contamination, this obviously affects the starting values and, consequently, the entire sequence of conditional variances. Because this effect is of order $1/n$ in case only a single outlier occurs, we ignore this here.

Define $v_t = e_t^2 - h_t^e$. Then

$$\begin{aligned} v_t &= \nu_t & &\text{for } t < \tau, \\ v_\tau &= e_\tau^2 - h_\tau^e = (\varepsilon_t^2 + \omega)^2 - h_\tau \\ &= \nu_\tau + \omega^2 + 2\omega\varepsilon_\tau, \\ v_{\tau+j} &= e_{\tau+j}^2 - h_{\tau+j}^e \\ &= \varepsilon_{t+j}^2 - h_{\tau+j} - \beta_1^{j-1}\alpha_1(\omega^2 + 2\omega\varepsilon_\tau) \\ &= \nu_{\tau+j} - \beta_1^{j-1}\alpha_1(\omega^2 + 2\omega\varepsilon_\tau) & &\text{for } j = 1, 2, \ldots. \end{aligned}$$

Recalling the definition of the lag polynomial $\pi(L)$, we can summarize the above as

$$v_t = (\omega^2 + 2\omega\varepsilon_\tau)\pi(L)I_t[t = \tau] + \nu_t. \tag{11}$$

Notice that (11) can be interpreted as a regression model for v_t, that is,

$$v_t = \xi x_t + \nu_t, \tag{12}$$

with

$$x_t = 0 \qquad \text{for } t < \tau,$$

$$x_\tau = 1$$

$$x_{\tau+k} = -\pi_k \qquad \text{for } k = 1, 2, \ldots,$$

and

$$\xi \equiv f(\omega) = \omega^2 + 2\omega\varepsilon_\tau = \omega^2 + 2\omega(e_\tau - \omega) = -\omega^2 + 2\omega e_\tau. \tag{13}$$

The parameter ξ in (12) can be estimated by least squares, that is,

$$\hat{\xi}(\tau) = \left(\sum_{t=\tau}^{n} x_t^2\right)^{-1} \left(\sum_{t=\tau}^{n} x_t v_t\right). \tag{14}$$

Solving (13) for ω, it follows that

$$\omega = e_\tau \pm \sqrt{e_\tau^2 - \xi}. \tag{15}$$

Hence, from the estimate $\hat{\xi}(\tau)$, the magnitude of the outlier at $t = \tau$ can be estimated as

$$\hat{\omega}(\tau) = \begin{cases} 0 & \text{if } e_\tau^2 - \hat{\xi}(\tau) < 0, \\ e_\tau - \sqrt{e_\tau^2 - \hat{\xi}(\tau)} & \text{if } e_\tau^2 - \hat{\xi}(\tau) > 0 \text{ and } e_\tau > 0, \\ e_\tau + \sqrt{e_\tau^2 - \hat{\xi}(\tau)} & \text{if } e_\tau^2 - \hat{\xi}(\tau) > 0 \text{ and } e_\tau < 0. \end{cases} \tag{16}$$

Notice that in (16) we assume that the outlier ω is of the same sign as the observed time series and, as $|\hat{\omega}(\tau)| < |e_\tau|$, is of the same sign as the realization of the core process ε_τ. Hence, the possibility of sign reversal is excluded.

To obtain a statistic which can be used to test for the presence of an outlier at $t = \tau$, which follows a similar strategy as in Chen and Liu (1993) and Tsay (1988), consider the t-statistic of $\hat{\omega}(\tau)$,

$$t_{\hat{\omega}(\tau)} = \frac{\hat{\omega}(\tau)}{\sigma_\nu \left(\frac{\partial f(\omega)}{\partial \omega} \left[\sum_{t=\tau}^{n} x_t^2\right] \frac{\partial f(\omega)}{\partial \omega}\right)^{-1/2}}, \tag{17}$$

where σ_ν is an estimate of the standard deviation of ν_t, and where it follows from (13) that

$$\frac{\partial f(\omega)}{\partial \omega} = -2\omega + 2e_\tau + 2\omega\frac{\partial e_\tau}{\partial \omega} = 2e_\tau.$$

One can then test for an outlier at time $t = \tau$ by comparing $t_{\hat{\omega}(\tau)}$ with an appropriate critical value.

Before we turn to determining critical values in the next subsection, we first proceed with a few remarks. As discussed in Chen and Liu (1993), the properties of the t-statistic (17) depend quite crucially on the estimate of the residual standard deviation σ_ν that is used. This is even more true here, as ν_t in (12) is not an i.i.d. process. Recalling that ν_t is defined as $\nu_t = h_t(z_t^2 - 1)$, it is immediately seen that ν_t is heteroskedastic, while the support of the distribution of ν_t changes every period. Some experimentation with different estimators for σ_ν suggests that both the leave-one-out estimator applied to the v_t series (which is one of the suggestions of Chen and Liu, 1993) and the usual sample standard deviation of the residuals $\hat{\nu}_t$ form the regression (12) perform reasonably well. It appears that it is not a good idea to use a heteroskedasticity-consistent standard error for $\hat{\omega}(\tau)$, that is, replacing the denominator of (17) with

$$\left(\sum_{t=\tau}^{n} x_t^2\right)^{-1} \left(\sum_{t=\tau}^{n} (\hat{\nu}_t x_t)^2\right) \left(\sum_{t=\tau}^{n} x_t^2\right)^{-1}. \tag{18}$$

For realistic values of the parameters α_1 and β_1 in the GARCH(1,1) model, the implied value of the regressor $x_{t+j} = -\pi_j = -\alpha_1\beta_1^{j-1}$ approaches zero quite rapidly as j increases, which makes the estimator (18) rather unreliable.

In contrast to the outlier detection statistics in the procedures of Tsay (1988) and Chen and Liu (1993), the t-statistic $t_{\hat{\omega}(\tau)}$ is *not* asymptotically standard normal distributed for fixed τ. Rather, the distribution depends on the parameters α_1 and β_1 in the GARCH(1,1) model. Intuitively, as either α_1 or β_1 increases, the kurtosis of ε_t increases is given by

$$\kappa_\varepsilon = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}, \tag{19}$$

under the additional assumption that z_t is i.i.d. standard normal. Note that κ_ε is always larger than the normal value of 3, and finite if $(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 < 1$. A larger value of κ_ε in turn increases the probability of large realizations of ε_t. The observation e_t then has to become even larger before it is to be viewed as an outlier.

In practice the location of possible outliers is of course unknown. An intuitively plausible test statistic is the maximum of the absolute values of the t-statistic over the entire sample, that is

$$t_{\max}(\hat{\omega}) \equiv \max_{1 \le \tau \le n} |t_{\hat{\omega}(\tau)}|. \tag{20}$$

The distribution of $t_{\max}(\hat{\omega})$ is highly non-standard and again depends on the parameters of the GARCH process which is specified for ε_t, as we will see below.

Finally, the outlier detection method for GARCH models consists of the following steps.

(i) Estimate a GARCH(1,1) model for the observed series e_t, obtain estimates of the conditional variance h_t^e and construct the series v_t as $v_t = e_t^2 - h_t^e$.

(ii) Obtain estimates $\hat{\omega}(\tau)$ for all possible $\tau = 1, \ldots, n$ using (14) and (15), and compute the test statistic $t_{\max}(\hat{\omega})$ from (20). If the value of the test statistic exceeds a pre-specified critical value C an outlier is detected at the observation for which the t-statistic of $\hat{\omega}$ is maximized (in absolute value), say $\hat{\tau}$.

(iii) Define the outlier corrected series e_t^* as $e_t^* = e_t$ for $t \neq \hat{\tau}$ and

$$e_{\hat{\tau}}^* = e_{\hat{\tau}} - \hat{\omega}. \tag{21}$$

(iv) Return to step (i) to estimate a GARCH(1,1) model for the series e_t^*.

The iterations terminate if the $t_{\max}(\hat{\omega})$ statistic no longer exceeds the critical value C.

3.2 Critical values

Application of the outlier detection method as described above requires an appropriate critical value C to determine the significance of the $t_{\max}(\hat{\omega})$ statistic. Table 1 displays selected percentiles of the distribution of the statistic under the null hypothesis of no outliers, for various choices of α_1 and β_1 and sample sizes $n = 250$ and 500. These percentiles are based on 5000 realizations of the GARCH(1,1) model (5) and (6), with z_t standard normally distributed. The parameters in the GARCH(1,1) model are estimated using the ML method as discussed in the previous section. The usual sample standard deviation of $\hat{\nu}_t$ is used to estimate σ_ν in (17).

- insert Table 1 about here -

It is seen that the percentiles of the distribution of $t_{\max}(\hat{\omega})$ depend rather heavily on the parameters in the GARCH model.

- insert Figures 7 and 8 about here -

Figures 7 and 8 contain scatterplots of the tabulated percentiles against the kurtosis of ε_t, as given by (19). It is seen that for $\kappa_\varepsilon < 4$, the relationship is close to linear. We find that the deviations from the linear relationship for higher values of the kurtosis can well be explained by the values of α_1 and β_1 in the GARCH(1,1) model. Hence, we consider the regression

$$\widehat{PC}_{\alpha,n} = b_0 + b_1\alpha_1 + b_2\beta_1 + b_3\kappa_\varepsilon + \eta_t, \tag{22}$$

where $\widehat{PC}_{\alpha,n}$ is the $100 \cdot \alpha$th percentile at sample size n, as shown in Table 1.

- insert Table 1 about here -

Least squares parameter estimates are shown in Table 1. The adjusted-R^2 values suggest that the fit of this simple equation is almost perfect. To determine the appropriate critical value in practice, in case one has approximately 250 or 500 observations, we therefore recommend that one estimates a GARCH(1,1) model, computes the implied kurtosis of e_t, $\hat{\kappa}_e$, and computes the appropriate critical values using (22). Before we turn to an investigation of the empirical performance of this strategy, we first give some further remarks.

- insert Table 3 about here -

To examine the effect of parameter estimation on the distribution of the $t_{\max}(\hat{\omega})$ statistic, Table 3 shows percentiles of this distribution obtained under the assumption that the parameters in the GARCH model are known. Comparing these percentiles with the ones given in Table 1 makes clear that the effect of estimating the parameters is not dramatic. The difference between the corresponding entries in the two tables usually is less than 0.5.

This suggests that an alternative method to obtain appropriate critical values is to consider the following bootstrap procedure. First, bootstrap samples are generated from the GARCH(1,1) model with the parameters that are obtained from fitting such a model to the observed series. Next, the outlier detection statistic is computed for the artificial series without estimating a GARCH model (but instead using the parameter estimates as obtained from the empirical time series). Upon repeating this B times, the resultant empirical distribution of the $t_{\max}(\hat{\omega})$ statistic can be used to judge the significance of the value of the test statistic that is found for the empirical time series. For example, the p-value of this statistic is the fraction of bootstrap samples for which the outlier detection statistic is larger than the empirical statistic. A similar procedure is suggested by Hotta and Tsay (1998).

To examine the usefulness of this procedure, we perform the following experiment. We generate series ε_t of length $n = 250$ from the GARCH(1,1) model (5)-(6) with $\omega = 0$, for the combinations of α_1 and β_1 considered in the previous section. A GARCH(1,1) model is estimated and the outlier detection statistic is computed. Using the parameter estimates, we generate $B = 499$ replications from a GARCH(1,1) model, and compute the outlier detection statistic without re-estimating the model. The p-value of the empirical statistic is then given by

$$\hat{p} = \frac{1}{B+1} \sum_{i=1}^{B} I[t_{\max}^i(\hat{\omega}) > t_{\max}(\hat{\omega})], \tag{23}$$

where $t_{\max}^i(\hat{\omega})$ denotes the test statistic for the i-th bootstrap sample. If this procedure is (approximately) valid, it should give correct size on average. Put differently, the p-values obtained according to (23) should be uniformly distributed on the interval [0,1].

- insert Figure 9 about here -

Figure 9 shows p-value discrepancy plots based on 500 replications. Based on these plots, we conclude that the bootstrap procedure works reasonably well.

3.3 Empirical power

To investigate the empirical power of the proposed outlier detection procedure, we apply the method to artificial series generated according to the procedure outlined in Section 2, with the following differences. We consider a more extensive set of values for the parameters α_1 and β_1 in (6). In addition to the combinations used before, we also set $(\alpha_1, \beta_1) = (0.1, 0.5)$, (0.1,0.7), (0.2,0.5) and (0.3,0.5). Furthermore, besides the case of a single outlier occurring halfway through the sample, we also consider the case where multiple outliers occur, at $\tau = \lfloor n/4 \rfloor$, $\lfloor n/2 \rfloor$ and $\lfloor 3n/4 \rfloor$. The magnitude of the outliers is varied among $\omega \in \{3, 4, 5\}$, while the outlier is restricted to have the same sign as ε_τ. All experiments are based on 1000 replications.

The results for the experiment involving a single outlier at $\tau = \lfloor n/2 \rfloor$ are shown in Table 4. The columns headed 'Detection' contain rejection frequencies of the null hypothesis of no outliers at the 5% significance level using critical values determined according to the bootstrap procedure outlined previously, with $B = 499$ bootstrap samples. The fraction of replications for which the outlier is located correctly is reported in columns headed ‘Location’. Note that these numbers do not take into account whether or not the null hypothesis is rejected by the test statistic. Corresponding figures conditional upon rejection are (even) closer to 1. Finally, columns headed ‘Magnitude’ report the mean and standard deviation of the estimate of the (absolute) magnitude of the outlier $|\hat{\omega}(\hat{\tau})|$, where again these numbers are computed using all replications, irrespective of whether or not the test statistic is significant the selected 5% significance level.

- insert Table 4 about here -

The results in Table 4 suggest the following conclusions. The empirical power of the test increases with an increase in the magnitude of the outlier. Second, when the sample size increases, the power gets smaller. This is intuitively plausible as it becomes more difficult to detect an outlier in larger samples as its effect becomes less apparent. Third, the method does not seem to face many problems with locating the outlier. Finally, the estimate of the value of ω seems rather precise, indicating that replacing the relevant observation with an outlier-corrected observation, using (21), is a sensible strategy.

- insert Tables 5 and 6 about here -

The results for the experiment involving multiple outliers are shown in Tables 5 and 6. In this experiment, the outlier detection procedure is applied in an iterative manner. That is, after estimating a GARCH(1,1) model, the $t_{\max}(\hat{\omega})$ is computed, and the observation for which the statistic is maximized is corrected, using the associated estimate of the magnitude of the outlier $\hat{\omega}$. Next, new estimates of the parameters in the GARCH(1,1) model are obtained from the corrected series, and so on. Table 5 contains the fraction of replications for which k outliers are detected, based on a 5% significance level, for $k = 1, 2, 3$. Note that the number of replications for which at least one outlier is found can be obtained by summing the fractions for $k = 1, 2$ and 3. For example, for the DGP with $\alpha_1 = 0.10$, $\beta_1 = 0.50$ and $n = 250$, at least one outlier is detected in 44% $(= (0.27 + 0.13 + 0.04) \times 100)$ of the replications. As in the previous experiment, this percentage increases as the magnitude of the outliers increases. Also, as ω gets larger, the frequency of detecting 2 or all 3 outliers increases.

The columns headed ‘Location’ in Table 6 contain fractions of replications for which the test statistic in the k-th iteration is maximized at one of the outlier locations for the experiment with $\omega = 3$. For larger values of ω, and/or conditional upon detection of an outlier these fractions all are very close to unity. As expected, the ability to locate an outlier correctly is largest in the first iteration. Note however that the fraction of correct location does not decrease all that much as k increases. Finally, the columns headed 'Magnitude' contain means and standard deviations of the estimate of the outlier at the k-th iteration (again these figures are based on all 1000 replications). The estimated magnitude of the outlier is largest in the first iteration, as could have been expected.

3.4 Extensions

Our proposed method remains valid if the ε_t's are shocks to a time series y_t, that, for example, is described by an ARMA(k,l) process subject to innovation outliers [IO's], that is,

$$\phi(L)y_t = \theta(L)\varepsilon_t, \tag{24}$$

$$y_t^* = y_t + \omega\frac{\theta(L)}{\phi(L)}I_t[t = \tau]), \tag{25}$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_k L^k$, $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_l L^l$, and where the conditional variance of ε_t evolves according to the GARCH(1,1) model (6). Notice that (24) and (25) can be combined to give

$$\phi(L)y_t^* = \theta(L)(\varepsilon_t + \omega I_t[t = \tau]), \tag{26}$$

from which it follows immediately that for the residuals of the observed time series $e_t = [\phi(L)/\theta(L)]y_t^*$ it still holds that $e_t = \varepsilon_t + \omega I_t[t = \tau]$. Hence, the results obtained above continue to hold in this case.

Notice that the procedures of Tsay (1988) and Chen and Liu (1993) applied to y_t do not work in the presence of GARCH. As $e_t = \varepsilon_t + \omega I_t[t = \tau] = z_t\sqrt{h_t} + \omega I_t[t = \tau]$, it is intuitively clear that observations for which the conditional variance h_t of the core process ε_t is larger are more likely to be identified as IO's than observations for which h_t is small. By using the transformation $v_t = e_t^2 - h_t$, we effectively (attempt to) eliminate this influence of h_t.

In contrast, the analysis of additive outliers [AO's] in an ARMA model with GARCH errors is much more cumbersome. For example, consider the case where the ARMA(k,l) process y_t is observed with occasional AO contamination,

$$y_t^* = y_t + \omega I_t[t = \tau]), \tag{27}$$

with y_t as defined in (24). Define the lag polynomial $\psi(L) = 1 - \psi_1 L - \psi_2 L^2 - \ldots$ as $\psi(L) = \phi(L)/(\theta(L)$. The residuals e_t can be obtained from $e_t = \psi(L)y_t^* = \psi(L)(y_t + \omega I_t[t = \tau]) = \varepsilon_t + \omega\psi(L)I_t[t = \tau]$, or in more detail

$$\begin{aligned} e_t &= \varepsilon_t && \text{for } t < \tau, \\ e_\tau &= \varepsilon_\tau + \omega, \\ e_{\tau+j} &= \varepsilon_{\tau+j} - \psi_j \omega && \text{for } j = 1, 2, \ldots. \end{aligned}$$

For the conditional variance computed with e_t according to (9) it then follows that $h_t^e = h_t$ for $t \leq \tau$, and

$$\begin{aligned} h_{\tau+1}^e &= \alpha_0 + \alpha_1 e_\tau^2 + \beta_1 h_\tau^e \\ &= \alpha_0 + \alpha_1(\varepsilon_\tau + \omega)^2 + \beta_1 h_\tau \\ &= h_{\tau+1} + \alpha_1(\omega^2 + 2\omega\varepsilon_\tau), \\ h_{\tau+2}^e &= \alpha_0 + \alpha_1 e_{\tau+1}^2 + \beta_1 h_{\tau+1}^e \\ &= \alpha_0 + \alpha_1(\varepsilon_{\tau+1} - \psi_1\omega)^2 + \beta_1(h_{\tau+1} + \alpha_1(\omega^2 + 2\omega\varepsilon_\tau)) \\ &= h_{\tau+2} + \alpha_1[\beta_1(\omega^2 + 2\omega\varepsilon_\tau) + (\psi_1^2\omega^2 - 2\psi_1\omega\varepsilon_{\tau+1})], \end{aligned}$$

and in general

$$h_{\tau+j}^e = h_{\tau+j} + \alpha_1 \sum_{i=0}^{j-1} \beta^{j-1-i}(\psi_i^2\omega^2 - 2\psi_i\omega\varepsilon_{\tau+i}), \qquad \text{for } j = 1, 2, \ldots, \tag{28}$$

where $\psi_0 \equiv -1$. Define again $v_t = e_t^2 - h_t^e$. Then

$$v_t = \nu_t \qquad \text{for } t < \tau, \tag{29}$$

$$v_\tau = e_\tau^2 - h_\tau^e = (\varepsilon_t^2 + \omega)^2 - h_\tau = \nu_\tau + \omega^2 + 2\omega\varepsilon_\tau, \tag{30}$$

$$\begin{aligned} v_{\tau+j} &= e_{\tau+j}^2 - h_{\tau+j}^e \\ &= \varepsilon_{t+j}^2 + \psi_j^2\omega^2 - 2\omega\psi_j\varepsilon_{\tau+j} - h_{\tau+j} - \alpha_1 \sum_{i=0}^{j-1} \beta^{j-1-i}(\psi_i^2\omega^2 - 2\psi_i\omega\varepsilon_{\tau+i}) \\ &= \nu_{\tau+j} + \psi_j^2\omega^2 - 2\omega\psi_j\varepsilon_{\tau+j} - \alpha_1 \sum_{i=0}^{j-1} \beta^{j-1-i}(\psi_i^2\omega^2 - 2\psi_i\omega\varepsilon_{\tau+i}) \qquad \text{for } j = 1, 2, \ldots. \end{aligned} \tag{31}$$

Even though (29)-(31) establishes the relationship between v_t and ν_t, it is not clear how this can be exploited to obtain an estimate of ω or a statistic which can be used to test whether $\omega = 0$.

The fact that we have to interpret the outliers as defined in (4)-(6) as IO's in a regression or time series model (and not as AO's) is not a drawback - given the use of GARCH models, it seems that the main interest lies in capturing the (second moment) properties of the shocks ε_t. Innovation outliers (or the outlier definition in (4)-(6)) then describe the situation where occasionally an aberrant shock occurs, which seems a natural possibility to consider in this case.

4 Application

In this section we apply our outlier detection and correction method to 10 years of weekly returns (1986-1995) on the stock markets of Amsterdam, Frankfurt, Paris, Hong Kong, Singapore and New York, which amounts to approximately 500 observations. We use weekly data for 1996-1998 to evaluate the out-of-sample forecast performance of conditional volatility with GARCH(1,1) models estimated on the series before and after outlier-correction.

- insert Table 7 about here -

Table 7 displays the estimated location of potential outliers and their magnitude for the 6 series under consideration, where a 5% significance level is used to asses the significance of the outlier detection statistic. Clearly, all series seem to have outliers around the October 1987 stock market crash. We find 1 or 2 outliers for 5 series, and 5 aberrant observations for Singapore.

- insert Table 8 about here -

In Table 8, we present parameter estimates for GARCH(1,1) models fitted to the series before and after outlier correction, where the locations and magnitudes of these outliers are given in Table 7. Comparing these estimates, one can notice that the differences between the parameter estimates can be as large as 0.200 (see for example $\hat{\beta}_1$ for Frankfurt and Hong Kong). This suggests that quite different forecasts can be obtained from the two estimated GARCH models. The estimated skewness and kurtosis in the final two columns of Table 8 indicate that the scaled residuals become 'closer to normal' after outlier correction.

- insert Table 9 about here -

Finally, in Table 9 we present some evidence that GARCH(1,1) models for outlier-corrected returns yield improved one-step-ahead forecasts of conditional volatility. Without exception, we find that outlier correction give significantly more accurate forecasts. For example, correcting only 2 obsevations (10/21/1987 and 10/28/1987) in the New York stock returns series results in a 25% reduction of the median squared prediction error.

5 Concluding remarks

In this paper we proposed an adaptation of the Chen and Liu (1993) method to detect and correct for outliers in the GARCH(1,1) model. We generated critical values for the relevant test statistic. Monte Carlo simulations suggested that our simple method works well in settings which are relevant for practical purposes. Our empirical results showed that correcting for only a few outliers yielded very different parameter estimates and significant improvements in out-of-sample forecast performance.

References

Bollerslev, T., 1986, Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* **31**, 307–327.

Bollerslev, T., 1988, On the correlation structure for the generalized autoregressive conditional heteroskedastic process, *Journal of Time Series Analysis* **9**, 121–131.

Bollerslev, T., R.Y. Chou and K.F. Kroner, 1992, ARCH modeling in finance: a review of the theory and empirical evidence, *Journal of Econometrics* **52**, 5–59.

Chen, C. and L-M. Liu, 1993, Joint estimation of model parameters and outlier effects in time series, *Journal of the American Statistical Association* **88**, 284–297.

Diebold, F.X. and R.S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* **13**, 253–263.

Fiorentini, G., G. Calzolari and L. Panatoni, 1996, Analytic derivatives and the computation of GARCH estimates, *Journal of Applied Econometrics* **11**, 399–417.

Franses, P.H. and H. Ghijsels, 1999, Additive outliers, GARCH and forecasting volatility, *International Journal of Forecasting* **15**, 1–9.

Friedman, B.M. and D.I. Laibson, 1989, Economic implications of extarordinary movements in stock prices (with comments and discussion), *Brookings Papers on Economic Activity* **20**, 137–189.

Hotta, L.K. and R.S. Tsay, 1998, Outliers in GARCH processes, unpublished manuscript, Graduate School of Business, University of Chicago.

Lee, S.-W. and B.E. Hansen, 1994, Asympotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Econometric Theory* **10**, 29–52.

Lumsdaine, R.L., 1996, Consistency and asympotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, *Econometrica* **64**, 575–596.

Sakata, S. and H. White, 1998, High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility, *Econometrica* **66**, 529–567.

Tsay, R.S., 1988, Outliers, level shifts, and variance changes in time series, *Journal of Forecasting* **7**, 1–20.

Table 1: Percentiles of distribution of $t_{\max}(\hat{\omega})$ statistic for detection of outliers in GARCH(1,1) models

		$n = 250$				$n = 500$				
α	β	80%	90%	95%	99%	80%	90%	95%	99%	K_ε
0.10	0.50	11.75	13.57	15.53	20.82	13.23	15.25	17.28	22.25	3.10
0.10	0.60	11.51	13.59	15.49	20.71	13.38	15.34	17.60	22.36	3.12
0.10	0.70	11.90	14.04	15.96	21.17	13.66	15.72	17.94	23.42	3.18
0.10	0.80	12.02	14.11	16.23	21.99	14.21	16.69	18.96	25.03	3.35
0.15	0.50	12.16	14.26	16.71	22.64	13.96	16.38	18.91	25.87	3.25
0.15	0.60	12.31	14.44	16.88	21.70	14.33	16.75	19.43	27.17	3.34
0.15	0.70	12.73	15.21	17.83	25.15	15.12	17.83	20.52	28.56	3.58
0.15	0.80	13.64	16.63	19.33	26.52	17.48	21.02	24.72	34.78	5.57
0.20	0.50	12.85	15.24	18.04	26.04	15.19	18.26	21.34	29.84	3.56
0.20	0.60	13.25	15.97	19.02	26.78	16.21	19.52	22.88	32.37	3.86
0.20	0.70	14.30	17.31	20.76	29.91	17.85	21.67	25.37	36.10	5.18
0.25	0.50	13.97	17.27	20.59	30.02	17.03	20.73	25.03	37.22	4.20
0.25	0.60	14.69	18.38	21.90	32.29	18.56	23.03	27.64	40.57	5.46
0.30	0.50	15.53	19.32	23.15	35.70	19.50	24.31	29.62	45.51	6.00

Percentiles of the distribution of the $t_{\max}(\hat{\omega})$ statistic (20) for detection of outliers in the GARCH(1,1) model (4)-(6). The table is based on 5000 replications, with parameters estimated with ML. The final column shows the kurtosis of ε_t as implied by the GARCH(1,1) model.

Table 2: Parameter estimates for equation (22), relating percentiles of the empirical distribution of $t_{\max}(\hat{\omega})$ statistic to α_1, β_1 and κ_ε

	$n = 250$				$n = 500$			
	80%	90%	95%	99%	80%	90%	95%	99%
b_0	8.12	8.07	8.34	8.22	6.31	5.58	5.30	1.82
	(0.33)	(0.51)	(0.62)	(1.97)	(0.39)	(0.55)	(0.64)	(1.19)
b_1	12.00	18.67	28.10	55.17	18.16	27.74	37.77	77.55
	(1.38)	(2.09)	(2.57)	(8.14)	(1.63)	(2.28)	(2.63)	(4.89)
b_2	1.13	1.99	2.92	3.68	3.32	4.39	4.51	7.36
	(0.54)	(0.83)	(1.02)	(3.22)	(0.64)	(0.90)	(1.04)	(1.93)
b_3	0.53	0.78	0.85	1.45	1.04	1.41	1.82	2.75
	(0.08)	(0.12)	(0.14)	(0.45)	(0.09)	(0.13)	(0.15)	(0.27)
$\overline{R}^2$	0.99	0.99	0.99	0.97	0.99	0.99	0.99	0.99
$\hat{\sigma}$	0.14	0.23	0.26	0.26	0.17	0.23	0.26	0.50

Estimates of the parameters in (22), based on the percentiles given in Table 1. Standard errors are given in parentheses.

Table 3: Percentiles of distribution of $t_{\max}(\hat{\omega})$ statistic for detection of outliers in GARCH(1,1) models

		$n = 250$				$n = 500$				
α	β	80%	90%	95%	99%	80%	90%	95%	99%	K_{ε}
0.10	0.50	11.84	13.73	15.77	21.09	13.37	15.36	17.44	22.66	3.10
0.10	0.60	11.88	13.82	15.92	21.50	13.55	15.55	17.69	23.32	3.12
0.10	0.70	11.95	13.91	16.13	21.80	13.77	15.91	18.14	23.92	3.18
0.10	0.80	12.27	14.46	16.60	22.38	14.43	16.86	19.11	25.47	3.35
0.15	0.50	12.21	14.51	16.82	22.68	14.15	16.67	19.27	25.68	3.25
0.15	0.60	12.54	14.72	17.12	23.32	14.61	17.25	19.79	27.03	3.34
0.15	0.70	12.85	15.24	17.89	24.89	15.36	18.34	21.03	28.64	3.58
0.15	0.80	13.60	16.51	19.78	26.71	17.51	21.07	24.74	34.09	5.57
0.20	0.50	13.02	15.58	18.28	26.14	15.52	18.50	21.70	29.90	3.56
0.20	0.60	13.42	16.08	19.04	27.33	16.30	19.68	23.26	32.75	3.86
0.20	0.70	14.11	17.34	20.74	28.54	17.99	21.97	26.34	36.81	5.18
0.25	0.50	13.95	17.10	20.49	31.54	17.21	21.20	25.67	36.22	4.20
0.25	0.60	14.77	18.19	21.81	32.81	18.80	23.13	27.86	40.00	5.46
0.30	0.50	15.22	19.00	23.43	36.23	19.58	24.38	29.70	43.53	6.00

Percentiles of the distribution of the $t_{\max}(\hat{\omega})$ statistic (20) for detection of outliers in the GARCH(1,1) model (4)-(6). The table is based on 5000 replications, with parameters assumed known. The final column shows the kurtosis of ε_t as implied by the GARCH(1,1) model.

Table 4: Properties of outlier detection procedure - single outlier case

α_1	β_1	ω	Detection 3	Detection 4	Detection 5	Location 3	Location 4	Location 5	Magnitude 3	Magnitude 4	Magnitude 5
							$n = 250$				
0.10	0.50		0.38	0.92	1.00	0.87	0.99	1.00	2.78	3.71	4.68
									(0.56)	(0.59)	(0.59)
0.10	0.60		0.41	0.94	0.99	0.88	0.99	1.00	2.80	3.74	4.71
									(0.55)	(0.56)	(0.57)
0.10	0.70		0.38	0.88	0.99	0.82	0.98	1.00	2.81	3.73	4.70
									(0.58)	(0.61)	(0.62)
0.10	0.80		0.34	0.83	0.96	0.85	0.99	1.00	2.77	3.70	4.67
									(0.58)	(0.60)	(0.60)
0.20	0.50		0.30	0.78	0.96	0.81	0.98	1.00	2.79	3.72	4.69
									(0.56)	(0.58)	(0.58)
0.20	0.60		0.29	0.72	0.91	0.79	0.96	0.99	2.83	3.73	4.69
									(0.59)	(0.59)	(0.59)
0.20	0.70		0.27	0.64	0.82	0.76	0.93	0.98	2.89	3.78	4.73
									(0.62)	(0.61)	(0.61)
0.30	0.50		0.20	0.52	0.76	0.69	0.90	0.96	2.88	3.75	4.70
									(0.69)	(0.65)	(0.63)
							$n = 500$				
0.10	0.50		0.30	0.86	1.00	0.76	0.99	1.00	2.84	3.74	4.72
									(0.55)	(0.58)	(0.59)
0.10	0.60		0.29	0.86	0.99	0.75	0.98	1.00	2.85	3.75	4.73
									(0.57)	(0.60)	(0.60)
0.10	0.70		0.27	0.84	0.99	0.76	0.99	1.00	2.86	3.78	4.76
									(0.58)	(0.61)	(0.61)
0.10	0.80		0.25	0.70	0.95	0.69	0.97	1.00	2.85	3.74	4.71
									(0.56)	(0.59)	(0.59)
0.20	0.50		0.20	0.63	0.94	0.67	0.94	0.99	2.91	3.79	4.77
									(0.59)	(0.60)	(0.60)
0.20	0.60		0.14	0.54	0.88	0.64	0.93	0.98	2.90	3.77	4.74
									(0.59)	(0.59)	(0.58)
0.20	0.70		0.12	0.40	0.70	0.56	0.85	0.95	3.00	3.82	4.76
									(0.69)	(0.63)	(0.60)
0.30	0.50		0.10	0.32	0.67	0.55	0.83	0.93	3.04	3.85	4.78
									(0.76)	(0.69)	(0.66)

The columns headed 'Detection' contain rejection frequencies of the null hypothesis of no outliers by the statistic $t_{\max}(\hat{\omega})$ as given in (20) at the 5% significance level. Columns headed 'Location' indicate the fraction of replications for which the test statistic is maximized at the observation for which the outlier occurs (irrespective of whether the test statistic is significant or not). Columns headed 'Magnitude' denote the mean and standard deviation (in parentheses) of the estimate of the magnitude of the outlier (based on all replications). Time series of n observations are generated according to the GARCH(1,1) model, where a single outlier of magnitude ω is added at $\tau = \lfloor n/2 \rfloor$. The table is based on 1000 replications.

Table 5: Rejection frequencies of outlier detection statistic - multiple outliers

			$\omega = 3$			$\omega = 4$			$\omega = 5$		
α_1	β_1	k	1	2	3	1	2	3	1	2	3
							$n = 250$				
0.10	0.50		0.27	0.13	0.04	0.10	0.05	0.68	0.02	0.00	0.94
0.10	0.60		0.31	0.13	0.04	0.11	0.03	0.71	0.02	0.01	0.94
0.10	0.70		0.29	0.13	0.04	0.09	0.07	0.63	0.03	0.01	0.92
0.10	0.80		0.27	0.10	0.05	0.12	0.07	0.58	0.06	0.03	0.84
0.20	0.50		0.22	0.09	0.03	0.15	0.08	0.46	0.08	0.02	0.77
0.20	0.60		0.21	0.07	0.04	0.13	0.07	0.43	0.08	0.03	0.74
0.20	0.70		0.19	0.07	0.06	0.16	0.06	0.38	0.10	0.06	0.62
0.30	0.50		0.15	0.03	0.04	0.15	0.04	0.28	0.11	0.03	0.54
							$n = 500$				
0.10	0.50		0.32	0.11	0.02	0.12	0.12	0.65	0.01	0.00	0.97
0.10	0.60		0.34	0.11	0.02	0.15	0.11	0.64	0.01	0.00	0.96
0.10	0.70		0.28	0.10	0.02	0.14	0.14	0.59	0.03	0.01	0.94
0.10	0.80		0.28	0.07	0.03	0.20	0.12	0.45	0.06	0.02	0.86
0.20	0.50		0.23	0.05	0.01	0.20	0.14	0.32	0.06	0.03	0.75
0.20	0.60		0.19	0.04	0.01	0.20	0.09	0.27	0.09	0.02	0.63
0.20	0.70		0.13	0.04	0.01	0.18	0.07	0.16	0.11	0.06	0.42
0.30	0.50		0.09	0.02	0.00	0.15	0.04	0.10	0.12	0.04	0.33

Fraction of replications for which k outliers are found by iterative detection-correction procedure, where the statistic $t_{\max}(\hat{\omega})$ as given in (20) is evaluated at the 5% significance level. Time series of n observations are generated according to the GARCH(1,1) model, where 3 outliers of magnitude $\omega = 3$ are added at $\tau = \lfloor n/4 \rfloor$, $\lfloor n/2 \rfloor$ and $\lfloor 3n/4 \rfloor$. The table is based on 1000 replications.

Table 6: Location and magnitude of outlier estimates - multiple outlier case

α_1	β_1	k	Location 1	Location 2	Location 3	Magnitude 1	Magnitude 2	Magnitude 3
					$n = 250$			
0.10	0.50		0.96	0.92	0.84	3.21	2.69	2.34
						(0.55)	(0.38)	(0.28)
0.10	0.60		0.97	0.92	0.84	3.24	2.71	2.35
						(0.53)	(0.39)	(0.29)
0.10	0.70		0.96	0.90	0.81	3.24	2.71	2.35
						(0.58)	(0.39)	(0.29)
0.10	0.80		0.95	0.91	0.82	3.24	2.70	2.36
						(0.58)	(0.38)	(0.29)
0.20	0.50		0.94	0.87	0.81	3.25	2.72	2.38
						(0.55)	(0.38)	(0.32)
0.20	0.60		0.90	0.88	0.82	3.24	2.73	2.40
						(0.56)	(0.40)	(0.32)
0.20	0.70		0.89	0.86	0.76	3.31	2.78	2.45
						(0.62)	(0.42)	(0.35)
0.30	0.50		0.83	0.81	0.80	3.28	2.78	2.47
						(0.66)	(0.44)	(0.38)
					$n = 500$			
0.10	0.50		0.93	0.84	0.69	3.28	2.74	2.41
						(0.56)	(0.39)	(0.29)
0.10	0.60		0.93	0.85	0.72	3.32	2.75	2.41
						(0.55)	(0.38)	(0.28)
0.10	0.70		0.93	0.86	0.72	3.30	2.77	2.43
						(0.57)	(0.39)	(0.29)
0.10	0.80		0.90	0.80	0.69	3.28	2.75	2.43
						(0.57)	(0.40)	(0.31)
0.20	0.50		0.85	0.81	0.68	3.31	2.79	2.44
						(0.58)	(0.39)	(0.31)
0.20	0.60		0.82	0.81	0.68	3.30	2.80	2.48
						(0.59)	(0.41)	(0.33)
0.20	0.70		0.76	0.71	0.63	3.35	2.86	2.54
						(0.66)	(0.47)	(0.37)
0.30	0.50		0.73	0.69	0.63	3.37	2.86	2.53
						(0.72)	(0.49)	(0.38)

Columns headed 'Location' indicate the fraction of replications for which the test statistic in the k-th iteration is maximized at the observation for which the outlier occurs (irrespective of whether the test statistic is significant or not). Columns headed 'Magnitude' denote the mean and standard deviation (in parentheses) of the estimate of the magnitude of the outlier in the k-th iteration (based on all replications). Time series of n observations are generated according to the GARCH(1,1) model, where 3 outliers of magnitude $\omega = 3$ are added at $\tau = \lfloor n/4 \rfloor$, $\lfloor n/2 \rfloor$ and $\lfloor 3n/4 \rfloor$. The table is based on 1000 replications.

Table 7: Detected outliers (at the 5% significance level) in weekly returns on stock indexes

Exchange	Date	$\hat{\omega}$
Amsterdam	10/28/87	−14.22
	10/21/87	−9.21
Frankfurt		
	10/28/87	−14.16
Paris		
	10/28/87	−16.07
Hong Kong		
	10/28/87	−26.11
	06/07/89	−15.85
Singapore	10/14/87	−14.28
	10/21/87	−15.40
	08/22/90	−9.32
	08/08/90	−9.51
	03/13/86	9.36
New York	10/21/87	−13.18
	10/28/87	−7.67

Location and magnitude of detected outliers in GARCH(1,1) models for weekly stock index returns, estimated for the sample January 1986-December 1995.

Table 8: Estimates of GARCH(1,1) models for weekly stock returns, before and after outlier correction

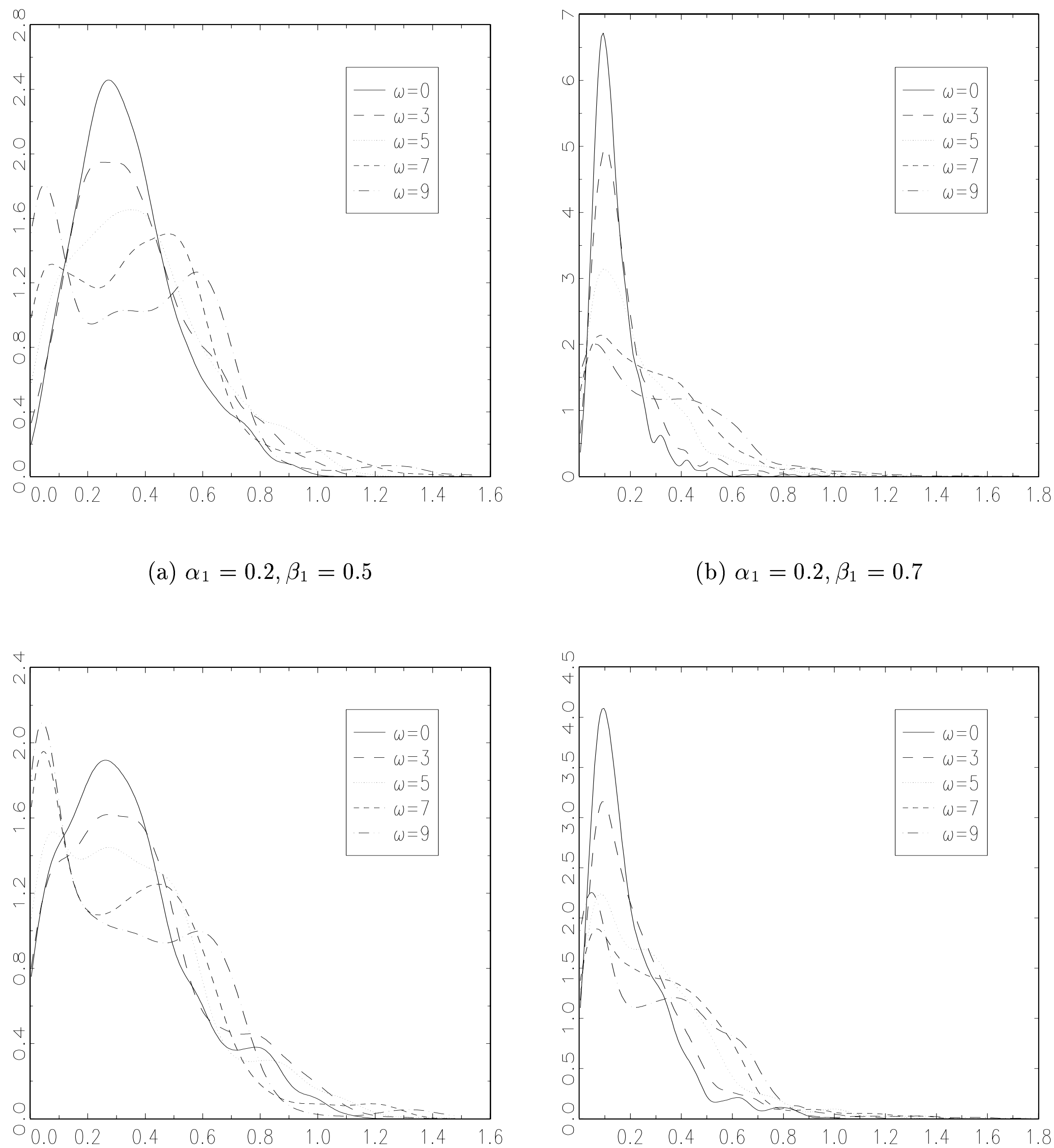
	μ	α_0	α_1	β_1	$SK_{\hat{z}}$	$K_{\hat{z}}$
Amsterdam						
Before	0.226	0.582	0.178	0.719	−0.88	5.61
	(0.095)	(0.147)	(0.027)	(0.045)		
After	0.198	0.310	0.094	0.840	−0.58	4.42
	(0.095)	(0.078)	(0.020)	(0.049)		
Frankfurt						
Before	0.227	1.488	0.146	0.646	−0.64	4.89
	(0.120)	(0.454)	(0.019)	(0.079)		
After	0.208	0.625	0.071	0.833	−0.50	4.24
	(0.120)	(0.191)	(0.013)	(0.090)		
Paris						
Before	0.187	0.940	0.163	0.717	−0.68	4.42
	(0.129)	(0.382)	(0.025)	(0.070)		
After	0.161	0.475	0.087	0.848	−0.55	3.91
	(0.118)	(0.248)	(0.036)	(0.107)		
Hong Kong						
Before	0.469	2.653	0.272	0.524	−0.94	5.39
	(0.135)	(0.725)	(0.028)	(0.083)		
After	0.449	1.315	0.132	0.744	−0.71	4.71
	(0.135)	(0.360)	(0.020)	(0.099)		
Singapore						
Before	0.273	1.180	0.091	0.731	−1.32	15.30
	(0.110)	(0.335)	(0.020)	(0.063)		
After	0.251	0.912	0.131	0.681	−0.11	4.23
	(0.110)	(0.259)	(0.024)	(0.061)		
New York						
Before	0.249	0.079	0.098	0.890	−1.15	7.92
	(0.081)	(0.032)	(0.019)	(0.025)		
After	0.234	0.033	0.045	0.944	−0.56	4.36
	(0.076)	(0.023)	(0.026)	(0.043)		

Estimates of GARCH(1,1) models for weekly stock index returns, $y_t = \mu + \varepsilon_t$, with $\varepsilon_t = z_t\sqrt{h_t}$ and $h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$. Models are estimated for the sample January 1986-December 1995. Standard errors based on the outer product of the gradient are given in parentheses. The final two columns contain the skewness and kurtosis of the standardized residuals $\hat{z}_t = \hat{\varepsilon}_t\hat{h}_t^{-1/2}$.

Table 9: Relative forecast performance of GARCH(1,1) models for weekly stock returns, after outlier correction against no correction

Exchange	MSPE	MedSPE	DM(Q)	MAE	MedAE	DM(A)
Amsterdam	0.75	0.80	4.77	0.88	0.90	6.77
Frankfurt	0.84	0.88	4.26	0.93	0.94	5.35
Paris	0.82	0.95	4.11	0.92	0.98	6.36
Hong Kong	0.71	1.00	3.12	0.90	1.00	3.92
Singapore	0.95	0.73	1.97	0.91	0.85	8.17
New York	0.66	0.75	6.05	0.82	0.86	9.63

Forecast performance of GARCH(1,1) models for weekly stock index returns, $y_t = \mu + \varepsilon_t$, with $\varepsilon_t = z_t\sqrt{h_t}$ and $h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$. Models are estimated for the sample January 1986-December 1995. One-step ahead forecasts are made for 1996-1998. True volatility is measured as $y_t - \overline{y}$, where $\overline{y}$ is the mean return over the estimation sample 1986-1995. Columns headed MSPE, MedSPE, MAE and MedAE contain ratios of the criteria for the model after outlier correction relative to the model before outlier correction. Columns headed DM(Q) and DM(A) contain the test statistic of Diebold and Mariano (1995), based on quadratic and absolute loss functions, respectively.



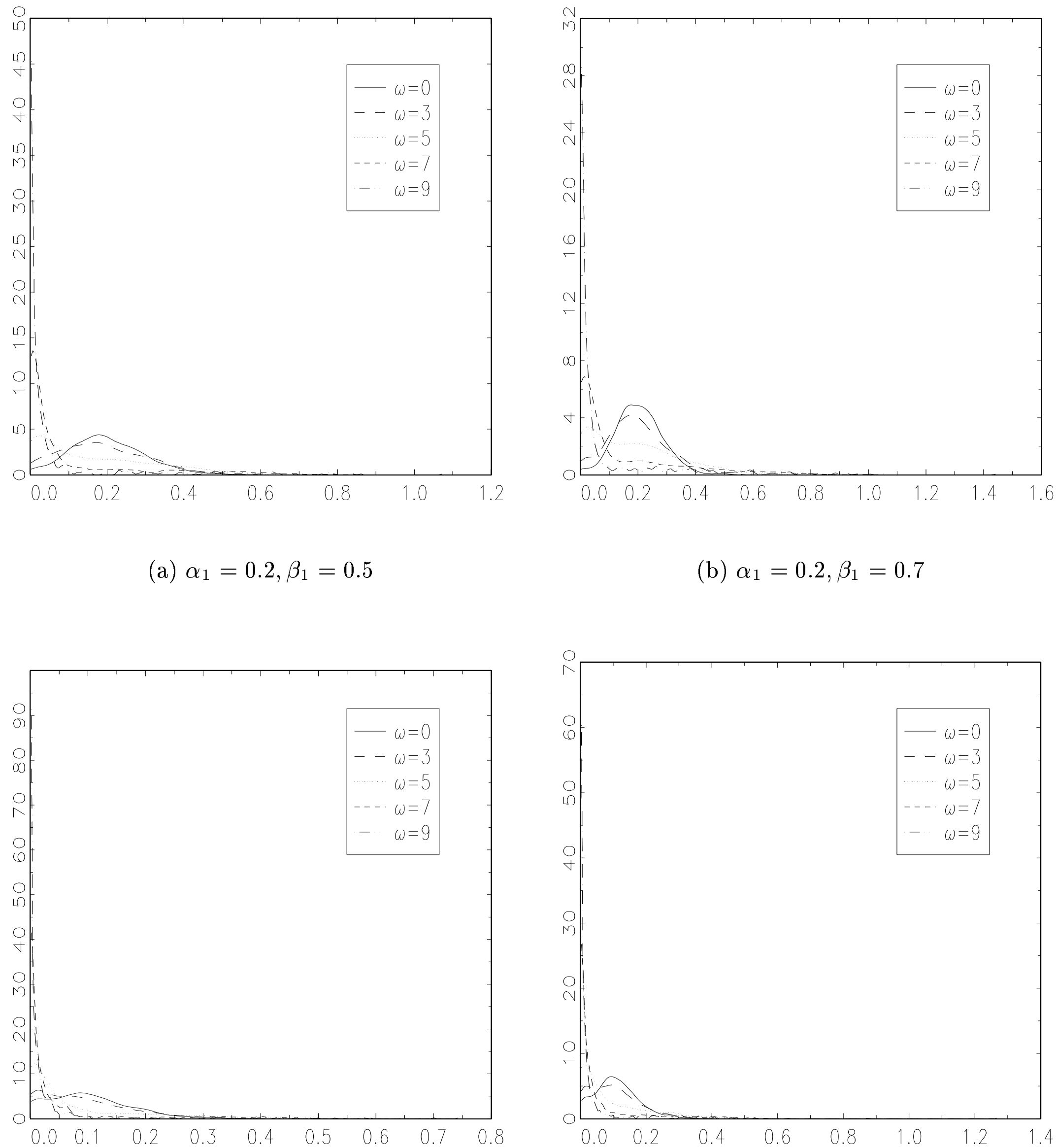
(a) $\alpha_1 = 0.2, \beta_1 = 0.5$

(b) $\alpha_1 = 0.2, \beta_1 = 0.7$

(c) $\alpha_1 = 0.1, \beta_1 = 0.6$

(d) $\alpha_1 = 0.1, \beta_1 = 0.8$

Figure 1: Empirical distributions of ML-estimates of α_0 for the contaminated GARCH(1,1) model (1)-(3) for series of $n = 250$ observations.



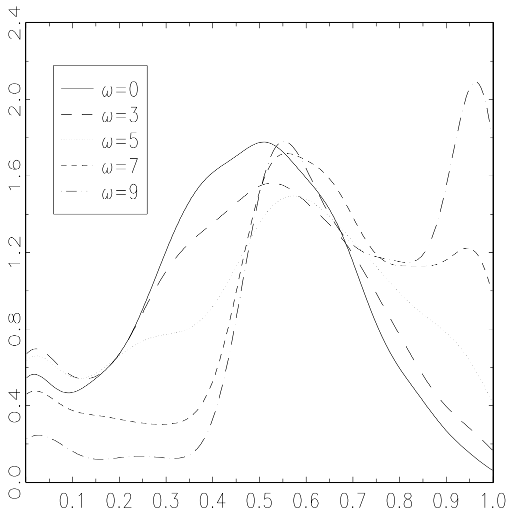
(a) $\alpha_1 = 0.2, \beta_1 = 0.5$

(b) $\alpha_1 = 0.2, \beta_1 = 0.7$

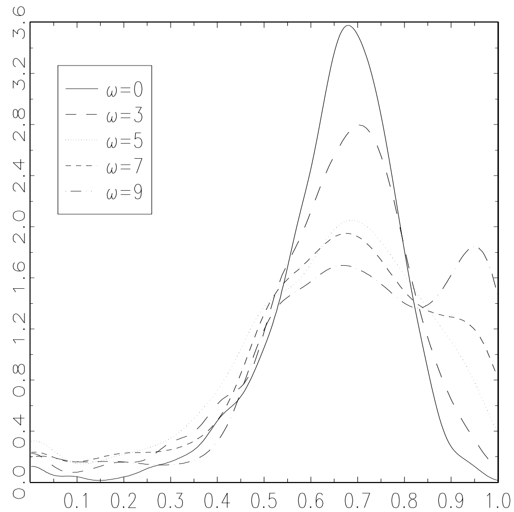
(c) $\alpha_1 = 0.1, \beta_1 = 0.6$

(d) $\alpha_1 = 0.1, \beta_1 = 0.8$

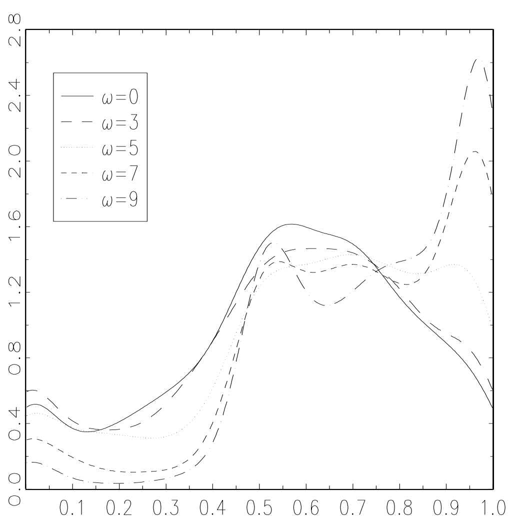
Figure 2: Empirical distributions of ML-estimates of α_1 for the contaminated GARCH(1,1) model (1)-(3) for series of $n = 250$ observations.



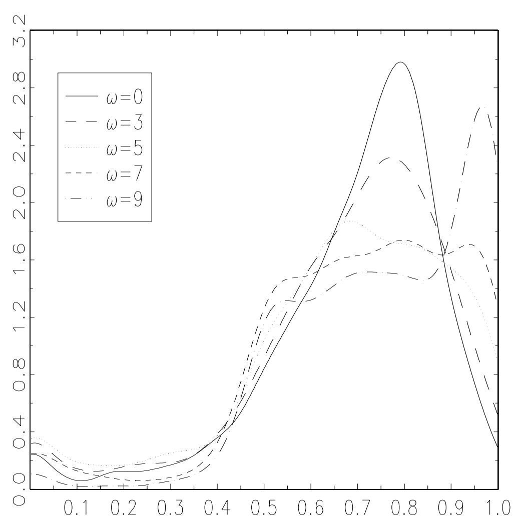
(a) $\alpha_1 = 0.2, \beta_1 = 0.5$



(b) $\alpha_1 = 0.2, \beta_1 = 0.7$

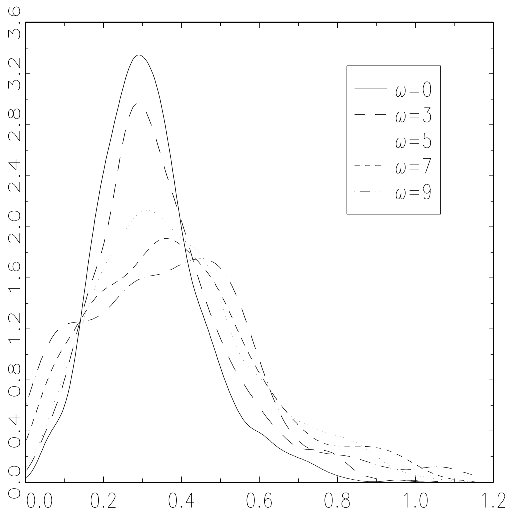


(c) $\alpha_1 = 0.1, \beta_1 = 0.6$

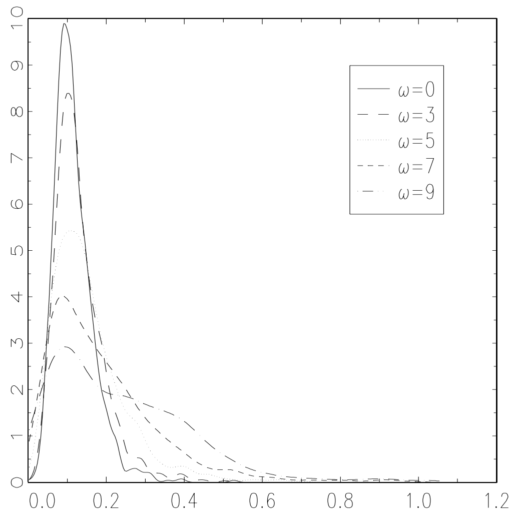


(d) $\alpha_1 = 0.1, \beta_1 = 0.8$

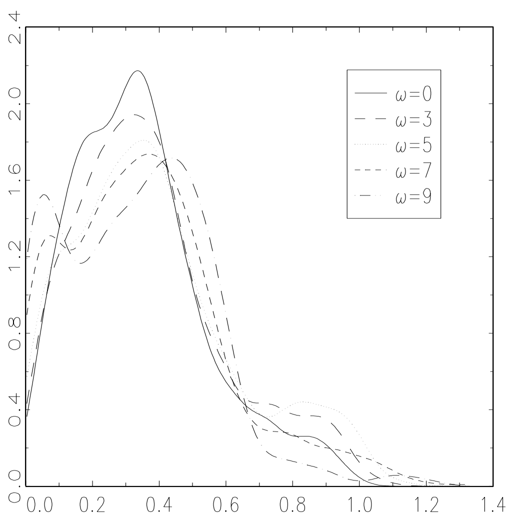
Figure 3: Empirical distributions of ML-estimates of β_1 for the contaminated GARCH(1,1) model (1)-(3) for series of $n = 250$ observations.



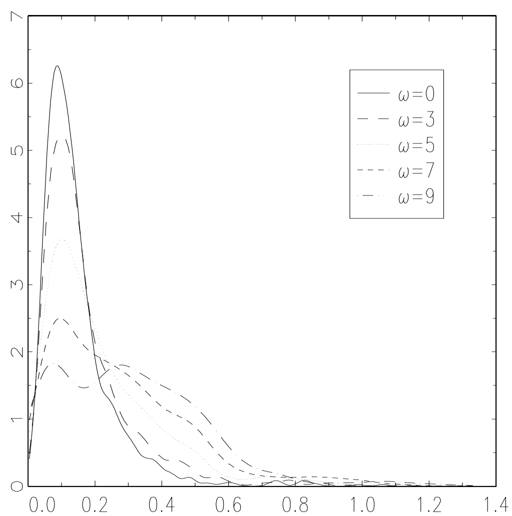
(a) $\alpha_1 = 0.2, \beta_1 = 0.5$



(b) $\alpha_1 = 0.2, \beta_1 = 0.7$

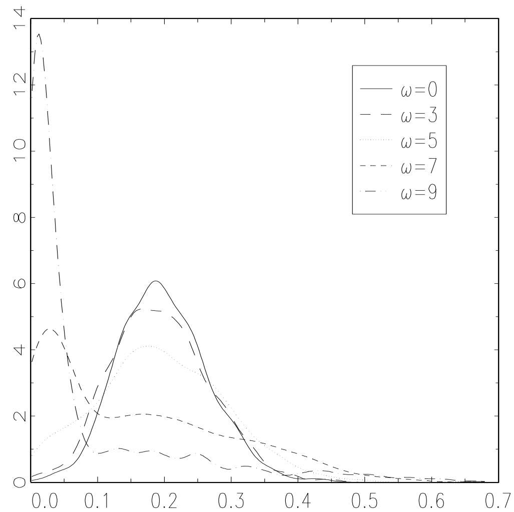


(c) $\alpha_1 = 0.1, \beta_1 = 0.6$

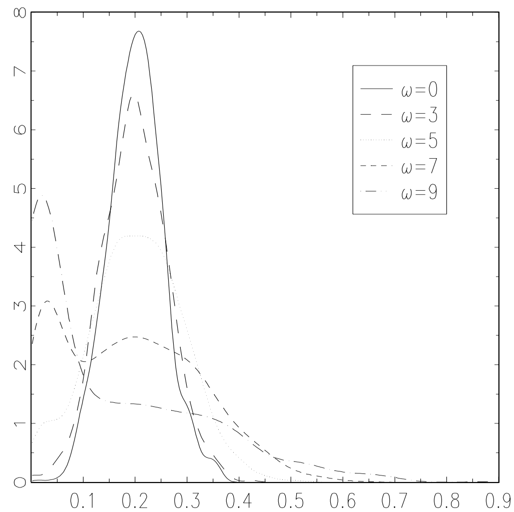


(d) $\alpha_1 = 0.1, \beta_1 = 0.8$

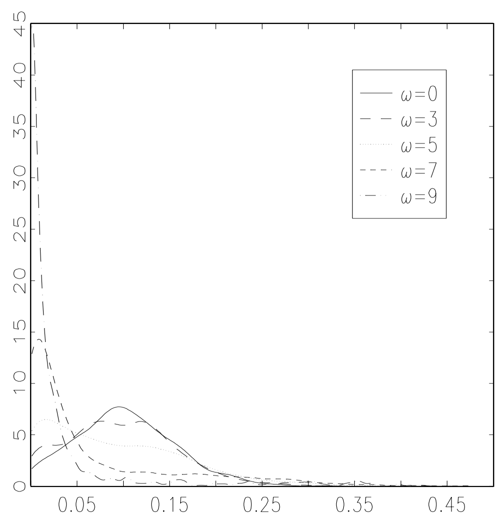
Figure 4: Empirical distributions of ML-estimates of α_0 for the contaminated GARCH(1,1) model (1)-(3) for series of $n = 500$ observations.



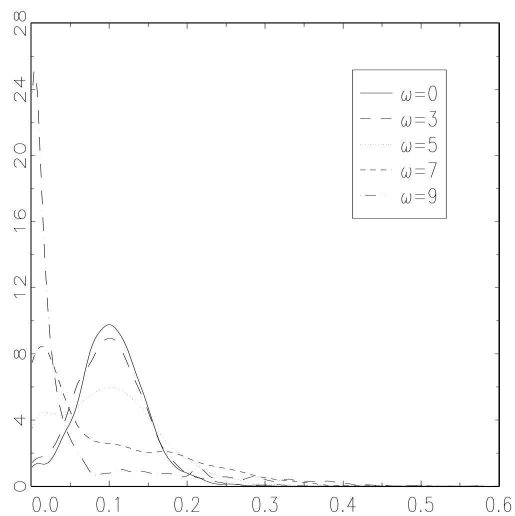
(a) $\alpha_1 = 0.2, \beta_1 = 0.5$



(b) $\alpha_1 = 0.2, \beta_1 = 0.7$

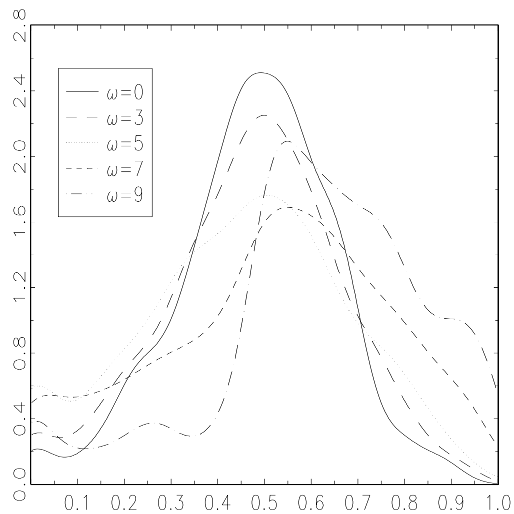


(c) $\alpha_1 = 0.1, \beta_1 = 0.6$

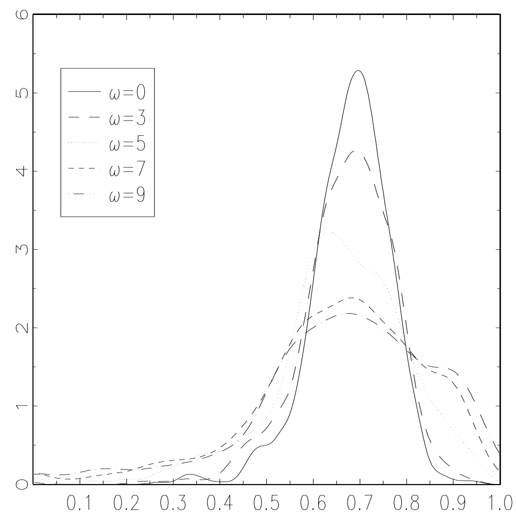


(d) $\alpha_1 = 0.1, \beta_1 = 0.8$

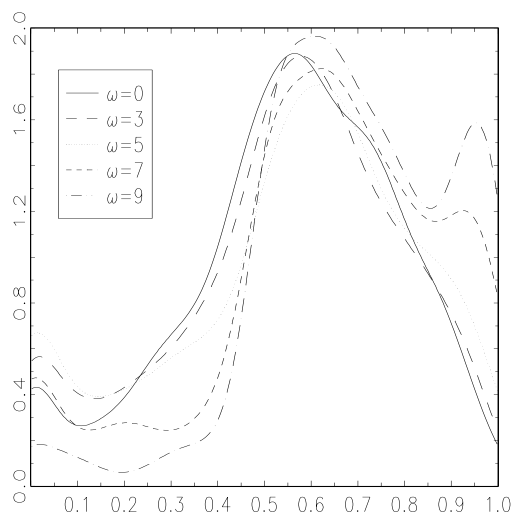
Figure 5: Empirical distributions of ML-estimates of α_1 for the contaminated GARCH(1,1) model (1)-(3) for series of $n = 500$ observations.



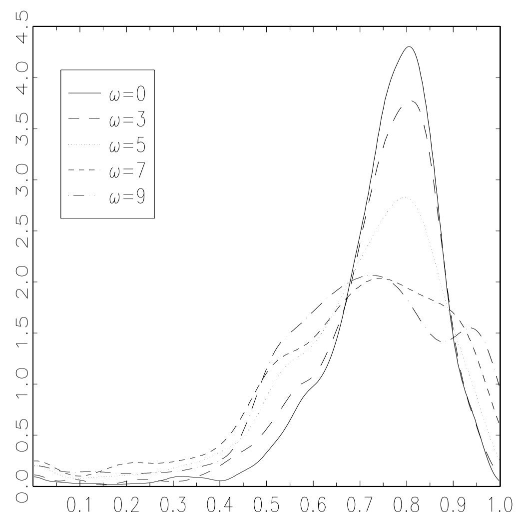
(a) $\alpha_1 = 0.2, \beta_1 = 0.5$



(b) $\alpha_1 = 0.2, \beta_1 = 0.7$

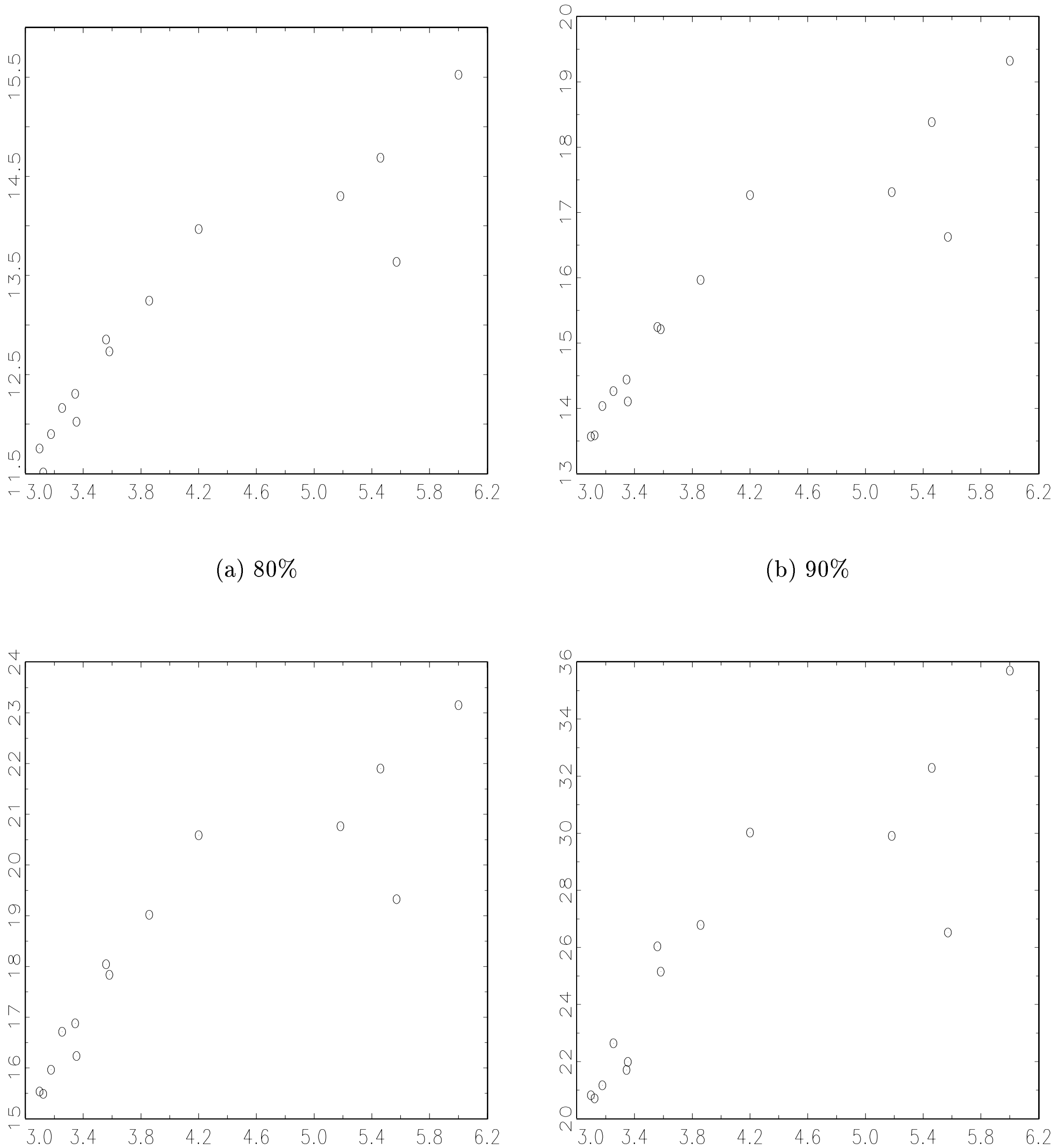


(c) $\alpha_1 = 0.1, \beta_1 = 0.6$



(d) $\alpha_1 = 0.1, \beta_1 = 0.8$

Figure 6: Empirical distributions of ML-estimates of β_1 for the contaminated GARCH(1,1) model (1)-(3) for series of $n = 500$ observations.



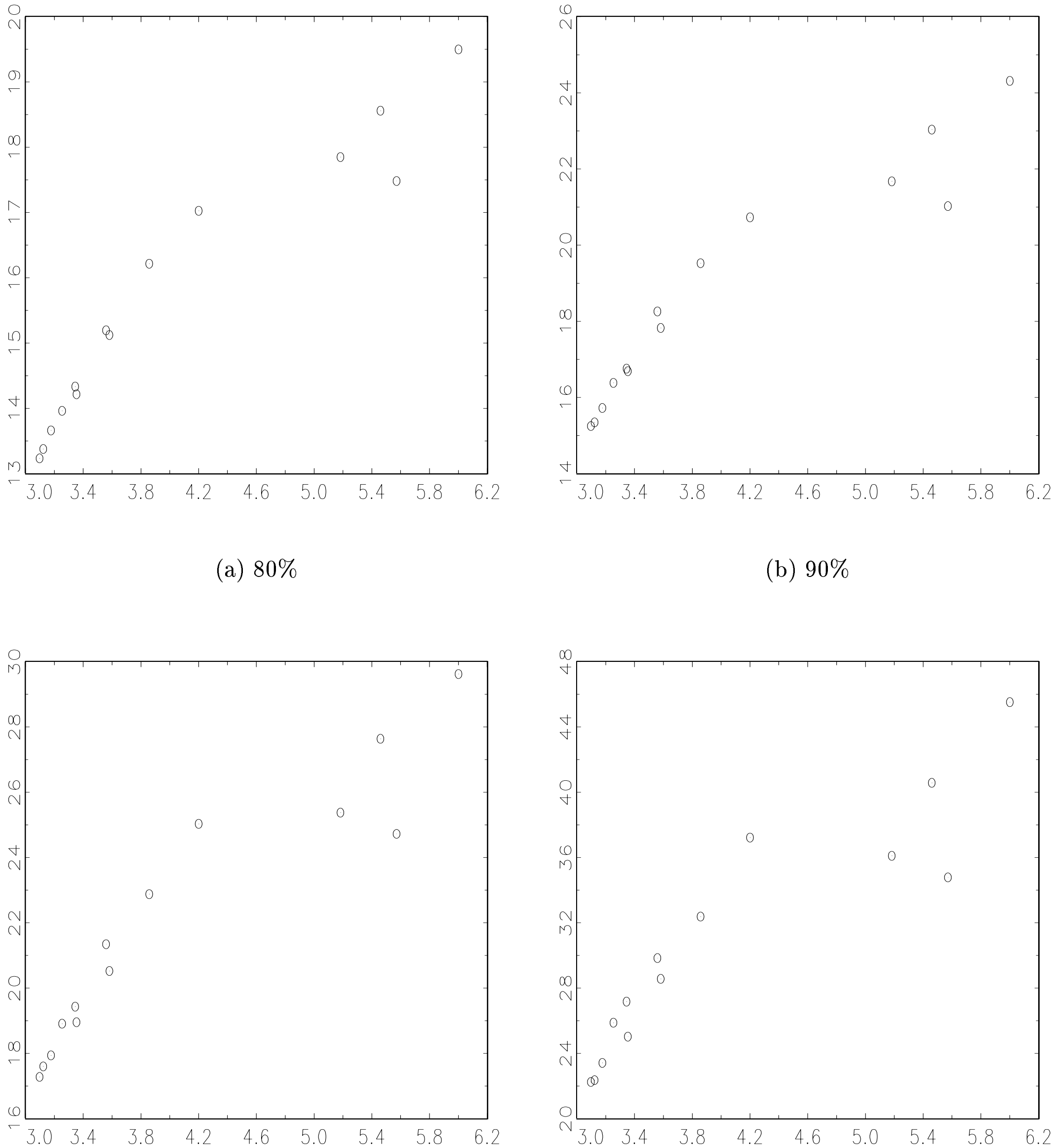
(a) 80%

(b) 90%

(c) 95%

(d) 99%

Figure 7: Scatterplot of percentiles of the empirical distribution of the $t_{\max}(\hat{\omega})$ statistic (20) for the detection of outliers in the GARCH(1,1) model (4)-(6) versus the kurtosis, for series of $n = 250$ observations.



(a) 80%

(b) 90%

(c) 95%

(d) 99%

Figure 8: Scatterplot of percentiles of the empirical distribution of the $t_{\max}(\hat{\omega})$ statistic (20) for the detection of outliers in the GARCH(1,1) model (4)-(6) versus the kurtosis, for series of $n = 500$ observations.

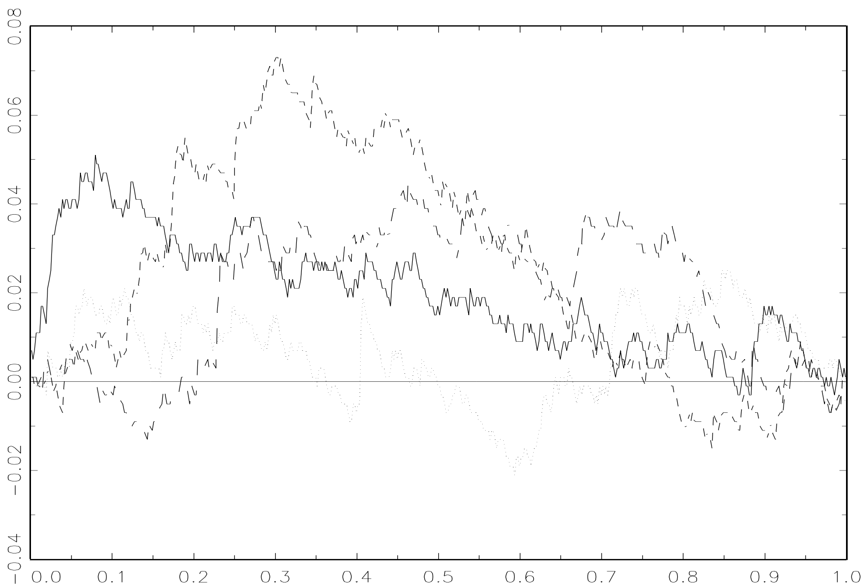


Figure 9: p-value discrepancy plots for bootstrap-with-fixed-parameters procedure, based on 500 replications of $n = 250$ observations, and $B = 499$ bootstrap samples. Series are generated according to the GARCH(1,1) model (4)-(6) with $\omega = 0$ and $(\alpha_1, \beta_1) = (0.2, 0.5)$ (solid line), (0.2,0.7) (long dashes), (0.1,0.6) (short dashes) and (0.1,0.8) (dotted line).