Estimation and Inference in ARCH Models in the Presence of Outliers

by

Allan W. Gregory and Jonathan J. Reeves *

Department of Economics
Queen's University
Kingston, Canada

January 2001

Abstract

In this paper we show the effects that outliers have on estimation and inference for ARCH models. We propose an empirically tractable solution to this problem by replacing outliers with their conditional expectations (optimal forecasts) in the likelihood function. This solution works well in both simulations and applications for a wide class of ARCH models. We demonstrate the accuracy of the procedure for parameter estimation, forecasting, and asset pricing. The empirical examples include U.S. interest rate, foreign exchange rate and stock index data. In addition, we offer a robust bootstrap test for outliers.

KEY WORDS: ARCH models; Foreign exchange rates; Interest rates; Outliers; Stock Prices.

*email: awg@qed.econ.queensu.ca and reevesj@qed.econ.queensu.ca. The authors would like to thank Russell Davidson for helpful suggestions. The first author acknowledges support of the Social Sciences and Humanities Research Council of Canada and the second author acknowledges support of the School of Graduate Studies and Research at Queen's University.

1 Introduction

The Autoregressive Conditional Heteroskedasticity (ARCH) class of models, introduced by Engle (1982), has become a core part of empirical finance. Indeed, citations of ARCH are too numerous to list (see Bollerslev et. al. (1992) for an excellent review.) These parsimonious models have been successful in capturing the volatility clustering so prevalent in financial data. Periods of high (low) volatility are autocorrelated and a variety of ARCH models have been developed and refined to account for some novel peculiarities in this persistence.

Despite the fact that periods of high volatility are serially correlated, there are occasions in which a singularly high or low observation for a series occurs, e.g. a financial crash, a merger announcement and so on, which does not appear to be part of the 'normal' data generating process (DGP). In most cases, there are too few of these outliers to model the process, and so outliers of this sort can not be predicted.

To motivate what we have in mind, we consider three prominent examples in the ARCH literature: Andersen and Lund (1997) who model the U.S. risk-free short-term interest rate; Glosten, Jagannathan and Runkle (1993) who model the U.S. risk premium; and West and Cho (1995) who model a number of foreign exchange rates. All three of these papers estimate ARCH models over the entire sample, without any consideration for possible outliers. We focus on these for illustrative purposes only, since most papers also neglect outliers.

Andersen and Lund (1997) estimate a Gaussian Level-EGARCH model for the U.S. risk-free short-term interest rate. However, from Figure 1, a large outlier can easily be identified corresponding to Black Monday (October 19, 1987). The second example is from monthly data, January 1952 to December 1998, and is similar to Glosten, Jagannathan and Runkle (1993). Stock prices are measured by the Standard & Poors 500 Index at close of the last trading day of each month and the risk premium is defined as the monthly return on the S & P 500 Index less the monthly return on the T-bill. From Figure 2, we see two major outliers: the October 1987 Crash and the stock market plunge of August 1998. Our final example uses foreign exchange data from West and Cho (1995) and is one we study in some detail. Figure 3 shows the weekly percentage change in the level of the exchange rate ($U.S./$Canadian) from March 7, 1973 to September 20, 1989. While not as clear-cut as the first 2 examples, one can identify at least 4 episodes that appear to be outliers (December 1976, March 1985, and the fall and rebound at the end of 1988).

In each case, there appears to be large departures from the 'normal' process generating the data and then an apparent return to this process. We would argue that at least at this frequency of data, the few outlier observations are not connected to the underlying process governing most of the observations. Nor in the case of multiple outliers is there any obvious way to establish an outlier process. The question then is what kind of effects such large outliers have on the estimates of ARCH models obtained from full sample information. In this paper we show quite dramatically that these outliers are high leverage observations, which result in substantially biased estimates and biased coverage probabilities for prediction intervals. We quantify the effect of these biases and propose a relatively simple solution that corrects both estimation and inference. The idea is to replace outlier observations by their conditional expectations (optimal forecasts) when building the likelihood function.

We would suggest that ARCH models are not designed to capture the extreme movements such as stock market crashes or foreign-exchange crises. Outliers of this magnitude in financial data are easily identifiable ex post. The procedure we propose is also retrospective in first identifying outliers then replacing these observations in the likelihood function with their expected values, conditional on information up to this time period. While our approach is conditional on first observing the data, the consequences of falsely identifying an outlier when it is not is in terms of efficiency loss which is negligible given the rather large sample sizes. However, estimation over the full sample, in the presence of an outlier, without a corrective method results in biased parameter estimates. For cases in which the researcher is uncertain, or wishes to test whether a particular observation or set of observations are outliers, we offer a robust bootstrap test based on the Hausman-Wu testing principle. Simulation evidence suggests this test has good size properties.

The Monte Carlo we conduct shows there are large biases in the parameter estimates and prediction intervals when outliers are ignored. Conditional expectations results in estimates and inferences that are almost as precise as the case in which no outliers are present. We also consider an option pricing example, which illustrates a rather large potential mis-pricing of an option in circumstances where outliers are ignored.

There are two procedures for handling outliers in ARCH models proposed in the literature. Sakata and White (1998) study the effect of outliers for a class of conditional dispersion models and propose the two-stage Hampel estimators and two-stage S-estimators which are resistant to the effect of outliers. Franses and Ghijsels (1999) design an iterative scheme to estimate the outlier effect for the GARCH(1,1) model. Applied researchers may find the procedure we outline much easier to implement and applicable to a wider class of ARCH models.

The paper is organized as follows. Section 2 sets up a general ARCH model and its quasi-maximum likelihood estimation. Section 3 describes our proposed method for ARCH estimation when outliers are present. Section 4 studies the foreign exchange example of West and Cho (1995). Section 5 provides Monte Carlo evidence demonstrating the effect outliers have on the estimation and inference of ARCH models and demonstrates the accuracy of our proposed estimation procedure. Section 6 discusses ARCH option pricing in the presence of outliers and Section 7 concludes.

2 The ARCH Model in a Simple Outlier Model

We first set up a very general ARCH model which encompasses most empirical specifications in the financial literature. Let Y_t , $t = 1, \ldots, \infty$ be a sequence of scalar random variables, $\{y_t : t = T - n + 1, \ldots, T\}$ be a realisation, and $\boldsymbol{x}_t = (y_{t-1}, y_{t-2}, \ldots, y_{T-n+1})$ denote the predetermined variables. We assume that y_t is governed by some ARCH model for the entire sample. When an outlier occurs instead of observing y_t, y_t^* is observed.

The conditional mean (μ_t) and variance function (Ω_t) are jointly parameterized by a finite dimensional vector $\boldsymbol{\theta}$:

$$\{\mu_t(\boldsymbol{x}_t, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\} \tag{1}$$

$$\{\Omega_t(\boldsymbol{x}_t, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\} \tag{2}$$

where $\boldsymbol{\Theta}$ is a compact subset of $\Re^p$ that has nonempty interior, and μ_t and Ω_t are known continuous functions of $\boldsymbol{x}_t$ and $\boldsymbol{\theta}$ that are twice continuously differentiable on the interior of $\boldsymbol{\Theta}$ for all $\boldsymbol{x}_t$.

Further, the first two conditional moments are correctly specified. For some $\boldsymbol{\theta}_0 \in$ *int* $\boldsymbol{\Theta}$

$$E(y_t|\boldsymbol{x}_t) = \mu_t(\boldsymbol{x}_t, \boldsymbol{\theta}_0)$$

$$V(y_t|\boldsymbol{x}_t) = \Omega_t(\boldsymbol{x}_t, \boldsymbol{\theta}_0) \ , \quad t = 1, 2, \ldots$$

with

$$y_t = \mu_t(\boldsymbol{x}_t, \boldsymbol{\theta}_0) + \varepsilon_t^0 \Omega_t^{\frac{1}{2}}(\boldsymbol{x}_t, \boldsymbol{\theta}_0) \tag{3}$$

$$E(\varepsilon_t^0|\boldsymbol{x}_t) = 0$$
$$E(\varepsilon_t^{0^2}|\boldsymbol{x}_t) = 1$$

Finally, the conditional variance satisfies:

$$0 < \Omega_t(\boldsymbol{x}_t, \boldsymbol{\theta}) < \infty \text{ for all } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Under the above conditions for the ARCH model and the technical assumptions in Appendix A of Bollerslev and Wooldridge (1992) - namely conditions A.1 (iii) - (vi), the quasi-maximum likelihood estimator is generally consistent for $\boldsymbol{\theta}_0$. (See Bollerslev and

Wooldridge (1992).) [2] Quasi-maximum likelihood estimation has become the standard estimation method for ARCH models. [3]

For observation t, the quasi-conditional log-likelihood (apart from a constant) is

$$l_t(\boldsymbol{\theta}; y_t, \boldsymbol{x}_t) = -\frac{1}{2} log|\Omega_t(\boldsymbol{x}_t, \boldsymbol{\theta})| - \frac{1}{2}(y_t - \mu_t(\boldsymbol{x}_t, \boldsymbol{\theta}))'\Omega_t^{-1}(\boldsymbol{x}_t, \boldsymbol{\theta})(y_t - \mu_t(\boldsymbol{x}_t, \boldsymbol{\theta}))$$

The quasi-maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is obtained by maximizing the quasi log-likelihood function

$$L_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta}) \tag{4}$$

We imagine a very simple outlier process. Let π be the probability of an outlier which is assumed to be independent of the process generating Y_t. This is a key assumption but seems defensible given the lack of a connection between outliers and the other observations. Denote y_t^{obs} as the value observed with probability structure:

$$y_t^{obs} = \begin{cases} y_t & \text{with probability } 1 - \pi \\ y_t^* & \text{with probability } \pi \end{cases}$$

This simple structure says that while the variable y_t is always determined by the ARCH model, equation 3, there are occasions for which we observe a contaminated or outlier value y_t^* whose magnitude and frequency is determined in some unknown but independent way. While we could, in principle build some time dependence in the outlier process, the empirical evidence suggests that this is unnecessary.

3 Estimation and Testing in the Presence of Outliers

The approach we propose is based on pre-identification of outliers prior to estimation. The identification of outliers in financial time series is often not difficult and can be done by simple graphical inspection. Outliers are often the result of an extreme market event

[2] Weaker conditions are sufficient for the GARCH(1,1) and IGARCH(1,1) models, see Lee and Hansen (1994) and Lumsdaine (1996).

[3] Andersen and Lund (1997) remarks are indicative of current practice; "in light of the quasi-maximum likelihood results of Bollerslev and Wooldridge (1992), we are more comfortable with the inference from the Gaussian version, although there is clear evidence of heavy tails in the conditional distributions. Indeed, if the $Student-t_v$ assumption is invalid, the maximum likelihood estimator is no longer consistent, while it retains consistency under the normality assumption, even in case of misspecification of the conditional density for ε_t."

such as a stock market crash and the date of the outlier is common knowledge as in Black Monday. As our three examples in the introduction show, it is easy to identify the extreme market events.

Once the outliers are identified, the quasi-maximum likelihood function can be maximized, but error terms that are affected by outliers are replaced with their expectation conditional on information up to the period before the outlier occurred. Thus, if there is an outlier at time t^* then $u_{t^*} = y_{t^*} - \mu_{t^*}(\boldsymbol{x}_{t^*}, \boldsymbol{\theta})$ is replaced with zero and $u_{t^*}^2$ with $\Omega_{t^*}(\boldsymbol{x}_{t^*}, \boldsymbol{\theta})$, since

$$E[u_{t^*}|\boldsymbol{x}_{t^*}] = 0 \tag{5}$$

and

$$E[u_{t^*}^2|\boldsymbol{x}_{t^*}] = \Omega_{t^*}(\boldsymbol{x}_{t^*}, \boldsymbol{\theta}) \tag{6}$$

While our approach is conditional on first observing the data and making some decisions about the existence of outliers; it should be kept in mind, the only consequence of mis-identifying an outlier when it is actually part of the normal DGP, is an efficiency loss. On the other hand, failure to remove an outlier results in biased parameter estimates.

We now discuss a robust test for outliers based on a non-parametric bootstrap test for situations in which one is uncertain of whether a particular observation or set of observations is having an influential impact on parameter estimation. The test is similar to a Hausman and Wu in which one compares a vector of contrasts.

Let $\widehat{\boldsymbol{\theta}}$ be the parameter estimates obtained from the entire sample and $\widetilde{\boldsymbol{\theta}}$ estimates obtained from the sample with suspected influential observations replaced with optimal forecasts. Let $\boldsymbol{V}(\widetilde{\boldsymbol{\theta}})$ be the Bollerslev and Wooldridge (1992) robust asymptotic variance matrix for $\widetilde{\boldsymbol{\theta}}$.

The null hypothesis H_0 is that the suspected influential observations are not influential and the alternative hypothesis H_A is that these observations are influential. The test statistic is:

$$\hat{\tau} = (\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}})'[\boldsymbol{V}(\widetilde{\boldsymbol{\theta}})]^{-1}(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}) \tag{7}$$

To estimate the bootstrap P value one takes the $\widetilde{\boldsymbol{\theta}}$ parameter estimates along with the estimated standardized empirical residuals. Theses residuals should be rescaled to have a mean of 0 and and a variance of 1 and should exclude the time periods for which there are suspected outliers. One then draws B bootstrap samples of which each is used to compute a bootstrap test statistic τ_j^* in exactly the same way as the real sample was used to compute $\hat{\tau}$. The bootstrap P value is estimated by

$$\hat{p}^*(\hat{\tau}) = \frac{1}{B}\sum_{j=1}^{B} I(\tau_j^* \geq \hat{\tau}),$$

where $I(.)$ is the indicator function.

We apply this test to our foreign exchange rate data from West and Cho (1995) and test whether the four outliers are influential. The test statistic is 10.801 with a P value of 0, from B=4999. Thus we can conclude that the four outliers, when not accounted for, are having a significant impact on parameter estimation.

Given the apparent soundly rejection of the null hypothesis of no influential observations, a rational question about the bootstrap test is its size properties. Are we observing a test that over-rejects. To examine this possibility we conduct a simple Monte Carlo experiment on test size. Unfortunately, due to the high computational demands of this experiment we are forced to limit ourselves to the case where the sample size is 350 and we use the above test for an outlier at observation 200. We do 1000 replications and 399 bootstrap draws for each replication.

The DGP we use is:

$$\text{(GARCH)} \quad \begin{array}{ll} y_t = u_t & t = 1, 2, \ldots, 350 \\ \omega_t^2 = 0.1 + 0.2u_{t-1}^2 + 0.7\omega_{t-1}^2 & , u_0^2 = 1 \text{ , } \omega_0^2 = 1 \\ u_t = \omega_t \varepsilon_t & \varepsilon_t \text{ i.i.d. standardized } t_5 \end{array}$$

which is fairly representative of many financial time series. In Table 1 the rejection frequencies for various significance levels are reported. The results are very favourable evidence for this bootstrap test statistic as the actual rejection frequencies are very close to the significance levels of the test.

Table 1. Rejection Frequencies for the Bootstrap Test with T = 350

Significance Level	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Rejection Frequency	0.013	0.022	0.033	0.041	0.054	0.064	0.069	0.079	0.090	0.095

4 Outliers in Exchange Rate Data

Our exchange rate data ($U.S./$Canadian) are Wednesday, New York noon bid rates, as published in the *The Federal Reserve Bulletin.* When Wednesday was a holiday we used Thursday data. We take the logarithmic difference of the exchange rate and then multiply by 100, as in West and Cho (1995) so that,

$$e_t = 100 * ln\left(\frac{\$U.S./\$Canadian \text{ in week t}}{\$U.S./\$Canadian \text{ in week t-1}}\right)$$

has the interpretation of percentage change in the level of the $U.S./$Canadian. As defined, we can interpret Figure 3 as the percentage change of the $U.S./$Canadian over the period from 7 March 1973 to 20 September 1989. The four largest movements in this time series occurred in December 1976, March 1985, and the fall and rebound at the end of 1988. Following West and Cho (1995), standard quasi-maximum likelihood estimation, without accounting for outliers, provides the following estimates:

$$e_t = u_t$$

$$\omega_t^2 = \underset{(0.0124)}{0.0283} + \underset{(0.1021)}{0.2254} u_{t-1}^2 + \underset{(0.1051)}{0.7063} \omega_{t-1}^2$$

$$u_t = \omega_t \varepsilon_t$$

Robust standard errors, as described in Bollerslev and Wooldridge (1992), are given below the parameter estimates.

Quasi-maximum likelihood estimation with outliers replaced with conditional expectations provides the following estimates:

$$e_t = u_t$$

$$\omega_t^2 = \underset{(0.0057)}{0.0149} + \underset{(0.0329)}{0.1324} u_{t-1}^2 + \underset{(0.0445)}{0.8183} \omega_{t-1}^2$$

$$u_t = \omega_t \varepsilon_t$$

After accounting for outliers, there is a significant change in all the parameters. The estimated unconditional variance falls from 0.4136 when outliers are not accounted for to 0.3029 when outliers are accounted for. The magnitude of the outliers, relative to the estimated unconditional standard deviation can be seen in Table 2.

Table 2. Scale of Outliers for $U.S/$C West and Cho (1995) Data Set

week end	obs.	value	$\hat{\omega}$	$\frac{\lvert value \rvert}{\hat{\omega}}$
01-12-76	e_{195}	-4.155	0.5504	7.55
27-02-85	e_{625}	-2.195	0.5504	3.99
02-11-88	e_{817}	-2.300	0.5504	4.18
23-11-88	e_{820}	2.551	0.5504	4.63

5 Some Monte Carlo Evidence

Monte Carlo experiments were conducted to analyse the effect of outliers on standard ARCH estimation and the accuracy of our proposed estimation procedure.[4]

Two main DGPs were used:

$$\begin{array}{lll} \text{Model 1.} & y_t = u_t + \psi I_t & t = 1, 2, \ldots, T \\ \text{(GARCH)} & \omega_t^2 = 0.1 + 0.2u_{t-1}^2 + 0.7\omega_{t-1}^2 & , u_0^2 = 1 \ , \omega_0^2 = 1 \\ & u_t = \omega_t \varepsilon_t & \varepsilon_t \text{ i.i.d. standardized } t_5 \end{array}$$

$$\begin{array}{lll} \text{Model 2.} & y_t = 0.5\omega_t + u_t + \psi I_t & t = 1, 2, \ldots, T \\ \text{(GARCH-M)} & \omega_t^2 = 0.1 + 0.2u_{t-1}^2 + 0.7\omega_{t-1}^2 & , u_0^2 = 1 \ , \omega_0^2 = 1 \\ & u_t = \omega_t \varepsilon_t & \varepsilon_t \text{ i.i.d. standardized } t_5 \end{array}$$

For both models:

$$I_t = \begin{array}{rl} 1 & \text{with probability } \frac{\pi}{2} \\ -1 & \text{with probability } \frac{\pi}{2} \\ 0 & \text{otherwise} \end{array}$$

We do experiments with sample sizes of 500 and 1000 which are quite typical of empirical finance applications. In each case we do not allow any outliers in the first 100 observations. We look at three cases: Case "No Outlier" is where we generate an ARCH process with $\psi = 0$ and do standard quasi-maximum likelihood estimation. Case "Outliers" is where we take the same ARCH process but now ψ is a positive number and we do standard quasi-maximum likelihood estimation. Case "Optimal Forecasts" is the same DGP as Case "Outliers" but now we do quasi-maximum likelihood estimation with outliers replaced with a conditional expectation. The values of ψ we use are 5, 7.5 and 10 which are realistic values associated with real financial time series. The values of π we use are $\frac{1}{200}$, $\frac{1}{400}$ and $\frac{1}{900}$ which again are realistic values associated with real financial time series.

[4] The experiments were performed using Ox version 2.20 and the simulation code is available upon request.

We do 4000 replications, but discard replications in which we do not get convergence for parameter estimates for all cases. We report the c % which is the percentage of successful replications out of 4000.

For each set of 4000 replications, we report the mean, standard deviation and root mean squared error of each parameter estimator for each case.[5]

We also report summary statistics on the coverage of the Gaussian 80 % and 95 % one step ahead prediction intervals. This is done by generating R=4000 true future values for each replication. These future values do not contain any outliers. The Gaussian prediction interval is calculated (L,U) for each case and the coverage is measured by $\frac{\#\{L \leq y^r_{T+1} \leq U\}}{R}$, where y^r_{T+1} $(r = 1, \ldots, R)$ are the true future values. The mean, standard deviation and root mean squared error of the estimated coverage for each case, over the number of successful replications, is reported. The results of these experiments are displayed in Tables 4 to 11 in Appendix B.

Our Monte Carlo work demonstrates the significant effect outliers can have on ARCH estimation and inference. And we see that our proposed solution of replacing outliers in the log likelihood function with their optimal forecast is very accurate, almost as accurate as the case where there are no outliers present. Thus replacing a few observations with their optimal forecast has negligible effect on the accuracy of the estimation. We would recommend that when in doubt over the influence of a few observations, it is safer to treat them as outliers and use the optimal forecast procedure.

The results for the GARCH(1,1) and GARCH(1,1)-M models are similar, as is going from a sample size of 500 to 1000. Four outliers occurring in a sample size of 1000 can lead the 80% prediction interval to give a coverage of close to 90% . And the coverage of the 95% prediction interval can get up to 97%. The mean, standard deviation and root mean squared errors for parameter estimates and prediction intervals are almost the same for the case where there are no outliers present and the case where one uses optimal forecasts in the log likelihood to correct for outliers. However, when outliers are not accounted for the accuracy of ones estimates can drop dramatically. The mean estimate on the constant in the skedastic function can triple. The co-efficient on the lagged variance falls. Whereas, the co-efficient on the squared residual can significantly increase when there are outliers of large magnitude. In addition there is a large increase in the standard deviation and root mean squared error of the estimates for the case where outliers are not accounted for. For the GARCH-M model the co-efficient in the conditional mean decreases when outliers are present. The decrease is larger when there are larger outliers occurring at higher frequency.

The convergence % drops as the magnitude of the outlier and the frequency of the outlier increase. For a relatively small outlier occurring at low frequency we get conver-

[5] These summary statistics are over the number of successful replications.

gence on almost every replication. The fact that outliers can affect the convergence of the estimation procedure is another reason why they should be properly accounted for.

6 ARCH Option Pricing in the Presence of Outliers

ARCH option pricing has become an important area of research in recent years. The majority of applications are based on simulation methods where an estimated ARCH process is simulated over the life of the option. Bollerslev and Mikkelsen (1996) find results that suggest that correctly modeling the volatility process of the underlying asset may be as important as the choice of approximate option valuation method when pricing long maturity contracts. And it is well known that deep-out-of-the-money long maturity options can be quite sensitive to the underlying volatility. Thus ignoring outliers in ARCH option pricing are in many cases likely to have unfortunate consequences. We demonstrate this with an empirical foreign exchange rate option pricing example.

A popular ARCH option pricing method is to follow Hull and White (1987), as in Noh, Engle and Kane (1994), Bollerslev and Mikkelsen (1996) and Engle, Kane and Noh (1997). This is the approach we follow in this example. The European call $C_{t,t+\tau}$ and put $P_{t,t+\tau}$ currency options τ periods ahead are valued as follows:

$$C_{t,t+\tau} = \frac{1}{N}\sum_{j=1}^{N} BS_j^C(S_t, K, \sigma_j^2, \tau) \tag{8}$$

$$P_{t,t+\tau} = \frac{1}{N}\sum_{j=1}^{N} BS_j^P(S_t, K, \sigma_j^2, \tau) \tag{9}$$

where $BS_j(.)$ represents the usual Black-Scholes option price formula which is:

$$BS_{t,t+\tau}^C = S_t e^{-r_f\tau}\Phi(d_1) - Ke^{-r\tau}\Phi(d_2) \tag{10}$$

$$BS_{t,t+\tau}^P = Ke^{-r\tau}\Phi(-d_2) - S_t e^{-r_f\tau}\Phi(-d_1) \tag{11}$$

where

$$d_1 = \frac{ln(S_t/K) + (r - r_f + \sigma^2/2)\tau}{\sigma\tau}$$

and

$$d_2 = d_1 - \sigma\tau$$

where $C_{t,t+\tau}, P_{t,t+\tau}$ are the Hull-White BS call and put option price forecasts at time t until the maturity date, S_t is the spot exchange rate (the value of one unit of the foreign currency in domestic currency), K is the exercise price, r is the home risk-free rate at time t, r_f is the foreign risk-free rate at time t, τ is the time to the maturity date, $\sigma_j^2 = (1/\tau)\sum_{i=1}^{\tau} \sigma_{t,t+i}^2$ is the volatility prediction at time t until the maturity date, which is generated by sampling randomly from the in-sample standardized residuals, for the particular ARCH model. Φ is the cumulative probability distribution function for a standard normal variable. N is the number of replications.

To produce an ARCH option price that is not affected by outliers, one replaces suspected outliers with their optimal forecasts and estimates the ARCH model, which is then used to generate the volatility prediction using the standardized empirical residuals, rescaled to have a mean of 0 and a variance of 1 and excluding time periods for which there are suspected outliers.

We applied these two procedures to price options on \$U.S/\$C using the West and Cho (1995) data set with the four outliers. We use the GARCH parameter estimates that were obtained in section 4. At the 20th of September 1989 our proxy for the US risk-free rate was .0773 and our proxy for the Canadian risk-free rate was .122 . At this date the spot exchange rate was 0.845 . Our time to maturity for our options are nine months, i.e. $\tau = 39$. All our option prices are based on $N = 1000$ replications. The superscript OC denotes outlier corrected.

Table 3. GARCH Simulated Nine Month Call and Put Option Prices

$\frac{K}{S_t}$	0.75	0.83	0.91	1	1.09	1.17	1.25
K	0.634	0.702	0.769	0.845	0.922	0.989	1.057
$C_{t,t+39}$	0.1837	0.1324	0.0910	0.0565	0.0336	0.0212	0.0134
$C^{oc}_{t,t+39}$	0.1794	0.1259	0.0828	0.0476	0.0254	0.0142	0.0078
$\frac{C_{t,t+39}}{C^{oc}_{t,t+39}}$	1.0243	1.0518	1.0988	1.1850	1.3203	1.4907	1.7298
$P_{t,t+39}$	0.0105	0.0233	0.0452	0.0823	0.1321	0.1829	0.2394
$P^{oc}_{t,t+39}$	0.0061	0.0168	0.0370	0.0735	0.1240	0.1760	0.2337
$\frac{P_{t,t+39}}{P^{oc}_{t,t+39}}$	1.7090	1.3876	1.2211	1.1199	1.0657	1.0396	1.0243

Table 3 demonstrates that we can get significantly different option prices when outliers are present and not accounted for, compared to when they are accounted for. For an at-the-money nine month call option the price increased by 18.5% when the outliers were not accounted for. And for an out-of-the-money nine month call option we have a 73 % increase in the price. We get similar large differences for at-the-money and out-of-the-money nine month put options.

7 Concluding Remarks

We have shown that ignoring outliers in ARCH estimation leads to biased parameter estimates and unreliable forecasts. Our solution of replacing outliers in the ARCH likelihood function with conditional expectations (optimal forecasts) leads to accurate estimation and inference. This solution is straightforward to implement, computationally fast and applicable to a wide class of ARCH models. Thus this procedure should be very useful for applications of ARCH estimation in the finance industry.

Appendix A

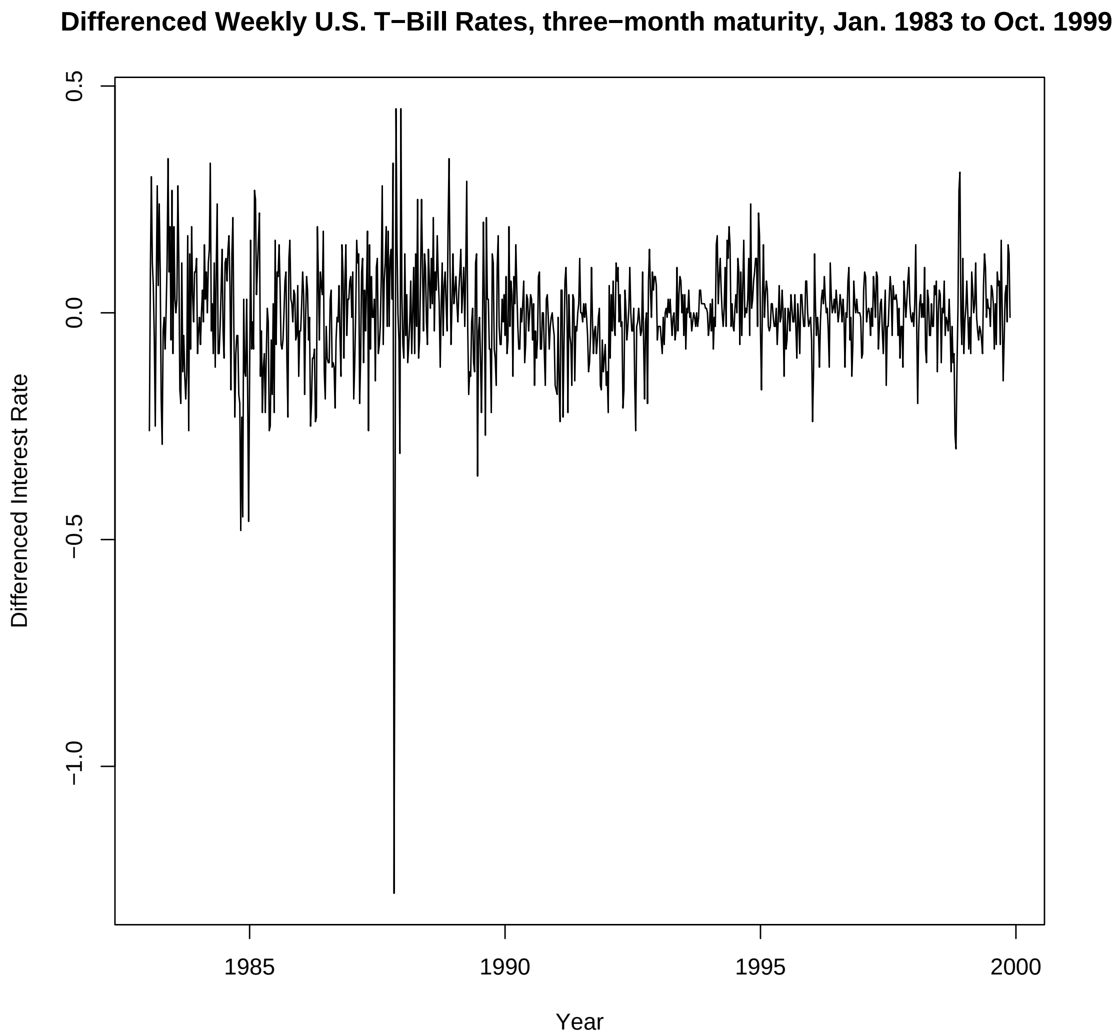

Figure 1:

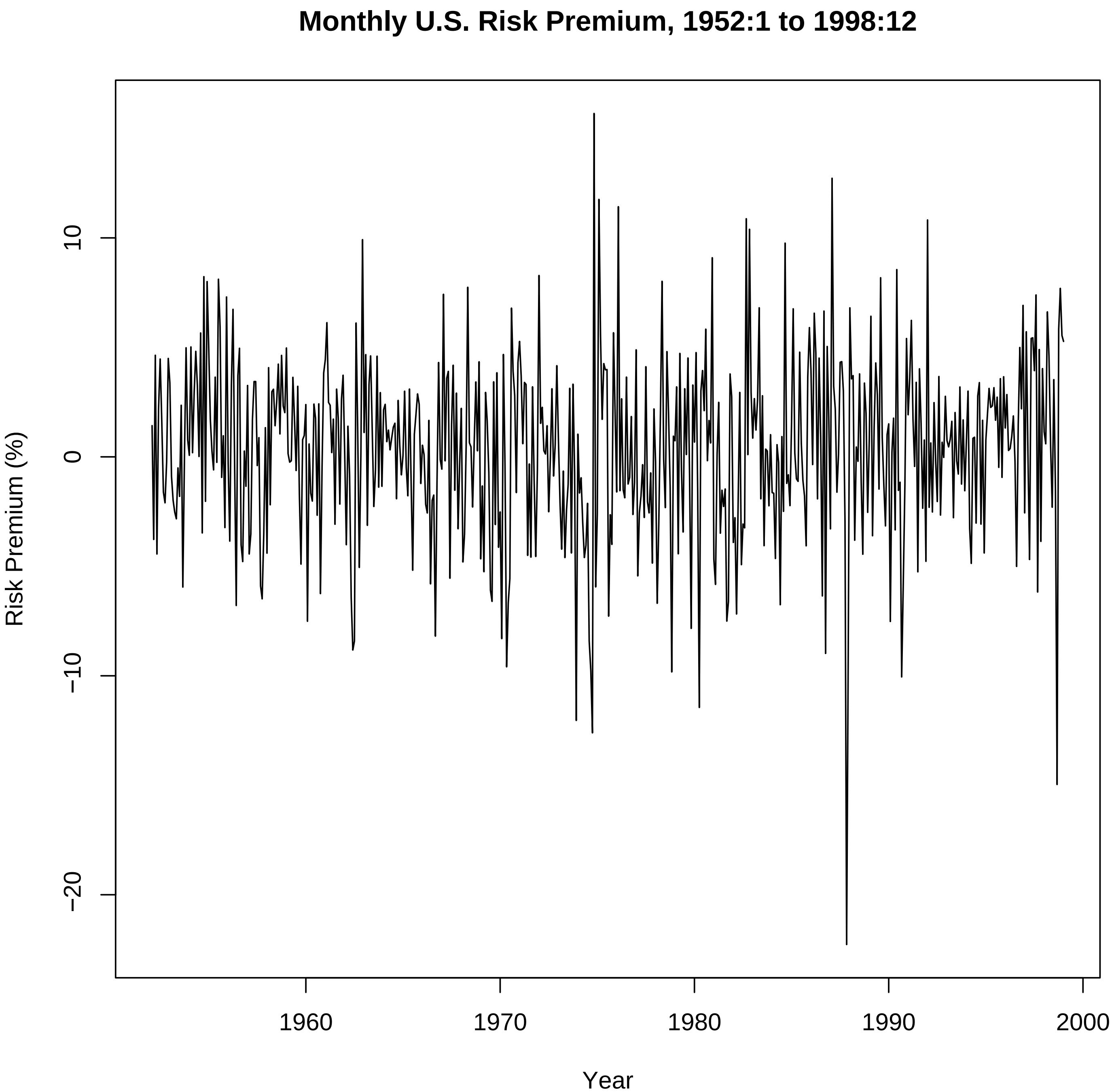
Monthly U.S. Risk Premium, 1952:1 to 1998:12
Risk Premium (%)
10
0
−10
−20
1960
1970
1980
1990
2000
Year

Figure 2:

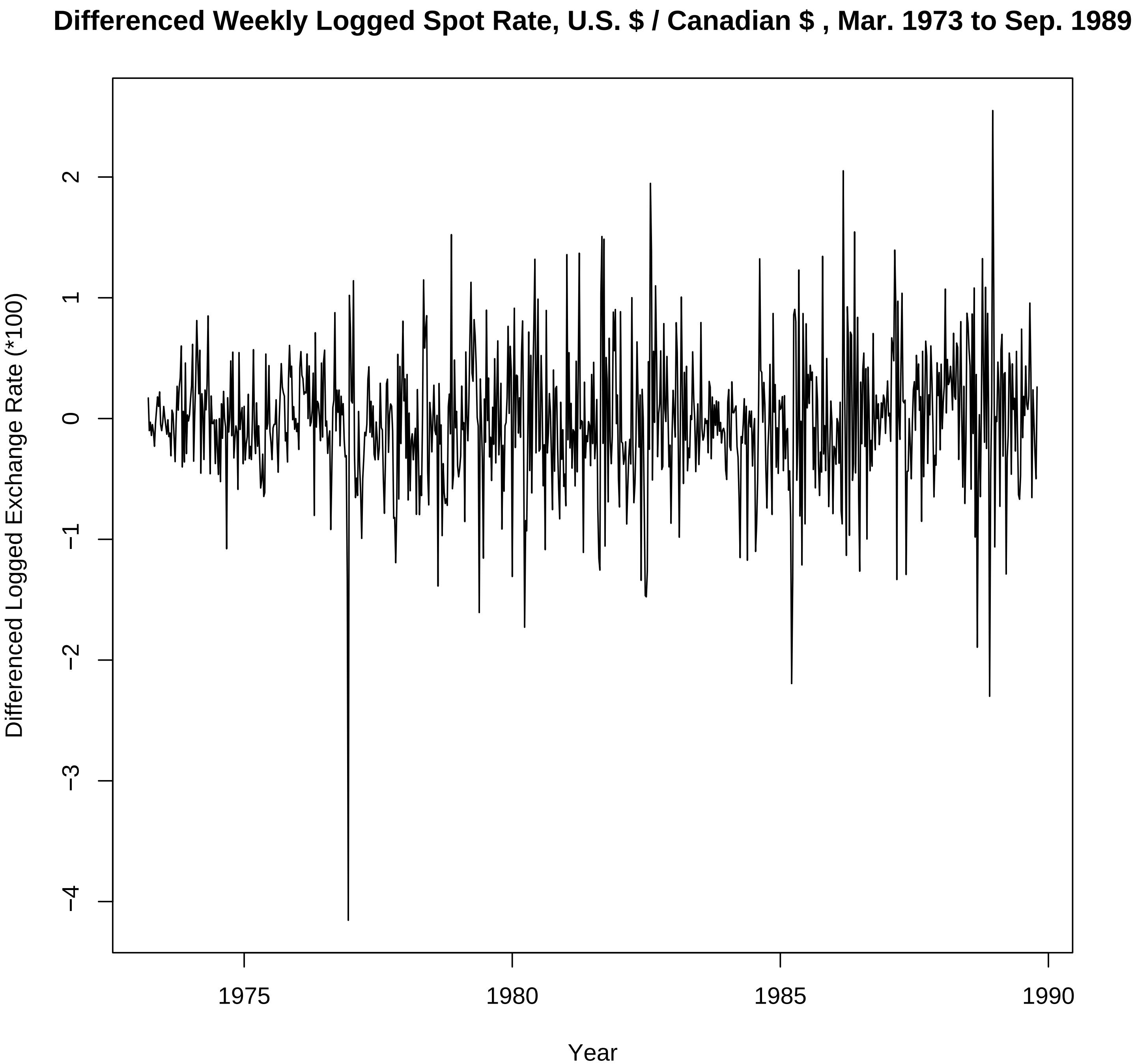
Differenced Weekly Logged Spot Rate, U.S. $ / Canadian $, Mar. 1973 to Sep. 1989
Differenced Logged Exchange Rate (*100)
2
1
0
−1
−2
−3
−4
1975
1980
1985
1990
Year

Figure 3:

Appendix B

Table 4. Estimates for Model 1 (GARCH(1,1)) with T = 500

Case				No Outlier			Outliers			Optimal Forecasts		
ψ	π		True	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
5	$\frac{1}{400}$	α_0	0.1	0.123	0.072	0.075	0.156	0.113	0.126	0.123	0.072	0.076
		α_1	0.2	0.211	0.154	0.155	0.210	0.167	0.167	0.210	0.154	0.154
		α_2	0.7	0.664	0.127	0.132	0.644	0.163	0.172	0.663	0.128	0.133
		PI	0.80	0.837	0.032	0.048	0.847	0.039	0.061	0.837	0.032	0.048
		PI	0.95	0.945	0.016	0.017	0.949	0.019	0.019	0.944	0.016	0.017
5	$\frac{1}{200}$	α_0	0.1	0.123	0.071	0.075	0.186	0.142	0.166	0.123	0.072	0.075
		α_1	0.2	0.211	0.155	0.155	0.207	0.175	0.175	0.210	0.154	0.154
		α_2	0.7	0.665	0.126	0.131	0.629	0.187	0.200	0.664	0.127	0.132
		PI	0.80	0.837	0.031	0.049	0.857	0.044	0.072	0.836	0.032	0.048
		PI	0.95	0.945	0.016	0.017	0.953	0.021	0.021	0.944	0.016	0.017
7.5	$\frac{1}{400}$	α_0	0.1	0.123	0.071	0.075	0.192	0.171	0.194	0.123	0.072	0.075
		α_1	0.2	0.212	0.155	0.156	0.226	0.148	0.151	0.211	0.155	0.155
		α_2	0.7	0.664	0.126	0.131	0.619	0.201	0.217	0.663	0.127	0.133
		PI	0.80	0.837	0.031	0.049	0.856	0.046	0.072	0.837	0.032	0.049
		PI	0.95	0.945	0.016	0.017	0.953	0.022	0.022	0.945	0.016	0.017
7.5	$\frac{1}{200}$	α_0	0.1	0.123	0.072	0.075	0.253	0.233	0.279	0.123	0.072	0.076
		α_1	0.2	0.212	0.156	0.157	0.232	0.184	0.186	0.212	0.156	0.156
		α_2	0.7	0.664	0.127	0.132	0.593	0.240	0.263	0.663	0.128	0.133
		PI	0.80	0.837	0.031	0.049	0.873	0.051	0.090	0.837	0.032	0.049
		PI	0.95	0.945	0.016	0.017	0.960	0.023	0.025	0.945	0.016	0.017
10	$\frac{1}{400}$	α_0	0.1	0.123	0.071	0.075	0.217	0.222	0.251	0.123	0.072	0.075
		α_1	0.2	0.213	0.157	0.158	0.261	0.220	0.228	0.213	0.157	0.157
		α_2	0.7	0.664	0.126	0.131	0.606	0.228	0.247	0.663	0.127	0.132
		PI	0.80	0.837	0.032	0.049	0.864	0.051	0.082	0.837	0.032	0.049
		PI	0.95	0.945	0.016	0.017	0.956	0.023	0.024	0.945	0.016	0.017
10	$\frac{1}{200}$	α_0	0.1	0.122	0.071	0.075	0.299	0.320	0.377	0.122	0.072	0.075
		α_1	0.2	0.214	0.161	0.162	0.291	0.301	0.315	0.213	0.161	0.161
		α_2	0.7	0.665	0.126	0.131	0.581	0.272	0.297	0.664	0.127	0.132
		PI	0.80	0.838	0.032	0.049	0.887	0.057	0.104	0.837	0.032	0.049
		PI	0.95	0.945	0.016	0.017	0.965	0.025	0.029	0.945	0.016	0.017

Table 5. Estimates for Model 1 (GARCH(1,1)) with T = 1000

Case				No Outlier			Outliers			Optimal Forecasts		
ψ	π		True	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
5	$\frac{1}{400}$	α_0	0.1	0.112	0.048	0.050	0.144	0.077	0.088	0.111	0.048	0.050
		α_1	0.2	0.203	0.065	0.065	0.200	0.076	0.076	0.203	0.065	0.065
		α_2	0.7	0.682	0.087	0.088	0.664	0.116	0.121	0.682	0.087	0.089
		PI	0.80	0.838	0.024	0.045	0.851	0.032	0.060	0.838	0.024	0.045
		PI	0.95	0.946	0.012	0.013	0.951	0.015	0.015	0.945	0.012	0.013
5	$\frac{1}{200}$	α_0	0.1	0.112	0.048	0.050	0.176	0.104	0.129	0.111	0.048	0.050
		α_1	0.2	0.203	0.065	0.065	0.194	0.086	0.086	0.202	0.065	0.065
		α_2	0.7	0.682	0.087	0.088	0.651	0.139	0.147	0.682	0.087	0.089
		PI	0.80	0.838	0.024	0.045	0.862	0.038	0.073	0.837	0.024	0.044
		PI	0.95	0.946	0.012	0.013	0.956	0.017	0.018	0.945	0.012	0.013
7.5	$\frac{1}{400}$	α_0	0.1	0.112	0.048	0.050	0.193	0.143	0.171	0.111	0.048	0.049
		α_1	0.2	0.204	0.065	0.065	0.215	0.107	0.108	0.203	0.065	0.065
		α_2	0.7	0.682	0.086	0.088	0.627	0.168	0.183	0.682	0.086	0.088
		PI	0.80	0.838	0.024	0.045	0.863	0.040	0.074	0.838	0.024	0.045
		PI	0.95	0.946	0.012	0.013	0.956	0.019	0.020	0.945	0.012	0.013
7.5	$\frac{1}{200}$	α_0	0.1	0.111	0.048	0.049	0.267	0.206	0.265	0.111	0.048	0.049
		α_1	0.2	0.204	0.065	0.065	0.214	0.138	0.139	0.203	0.065	0.065
		α_2	0.7	0.683	0.086	0.088	0.600	0.213	0.236	0.682	0.087	0.088
		PI	0.80	0.838	0.023	0.045	0.882	0.045	0.094	0.838	0.024	0.045
		PI	0.95	0.946	0.012	0.012	0.964	0.020	0.024	0.945	0.012	0.013
10	$\frac{1}{900}$	α_0	0.1	0.112	0.048	0.050	0.169	0.133	0.150	0.112	0.048	0.050
		α_1	0.2	0.204	0.065	0.065	0.226	0.120	0.123	0.203	0.065	0.065
		α_2	0.7	0.683	0.086	0.088	0.639	0.160	0.171	0.682	0.086	0.088
		PI	0.80	0.838	0.024	0.045	0.856	0.038	0.068	0.838	0.024	0.045
		PI	0.95	0.946	0.012	0.013	0.953	0.018	0.018	0.946	0.012	0.013
10	$\frac{1}{400}$	α_0	0.1	0.111	0.047	0.049	0.245	0.209	0.254	0.111	0.047	0.049
		α_1	0.2	0.204	0.065	0.065	0.242	0.161	0.166	0.204	0.065	0.065
		α_2	0.7	0.683	0.086	0.087	0.598	0.209	0.232	0.683	0.086	0.088
		PI	0.80	0.838	0.024	0.045	0.874	0.047	0.088	0.838	0.024	0.045
		PI	0.95	0.946	0.012	0.012	0.960	0.022	0.024	0.946	0.012	0.013
10	$\frac{1}{200}$	α_0	0.1	0.111	0.048	0.049	0.359	0.319	0.410	0.111	0.048	0.049
		α_1	0.2	0.206	0.065	0.065	0.251	0.223	0.229	0.205	0.065	0.065
		α_2	0.7	0.682	0.085	0.087	0.571	0.263	0.293	0.682	0.086	0.088
		PI	0.80	0.839	0.024	0.045	0.901	0.051	0.113	0.838	0.024	0.045
		PI	0.95	0.946	0.012	0.012	0.971	0.022	0.030	0.946	0.012	0.013

Table 6. Estimates for Model 2 (GARCH(1,1)-M) with T = 500

Case				No Outlier			Outliers			Optimal Forecasts		
ψ	π		True	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
5	$\frac{1}{400}$	ϕ	0.5	0.507	0.053	0.054	0.489	0.056	0.058	0.507	0.053	0.054
		α_0	0.1	0.119	0.064	0.067	0.150	0.100	0.112	0.119	0.064	0.067
		α_1	0.2	0.208	0.087	0.088	0.208	0.100	0.100	0.208	0.087	0.087
		α_2	0.7	0.669	0.113	0.117	0.650	0.145	0.153	0.669	0.114	0.118
		PI	0.80	0.828	0.044	0.052	0.839	0.050	0.064	0.828	0.044	0.052
		PI	0.95	0.941	0.026	0.028	0.946	0.028	0.028	0.941	0.027	0.028
5	$\frac{1}{200}$	ϕ	0.5	0.506	0.053	0.053	0.473	0.058	0.064	0.508	0.053	0.054
		α_0	0.1	0.119	0.064	0.066	0.179	0.126	0.149	0.119	0.064	0.067
		α_1	0.2	0.209	0.087	0.088	0.206	0.111	0.111	0.208	0.087	0.087
		α_2	0.7	0.669	0.112	0.116	0.637	0.167	0.178	0.669	0.113	0.117
		PI	0.80	0.828	0.044	0.052	0.850	0.053	0.073	0.828	0.044	0.052
		PI	0.95	0.941	0.026	0.028	0.951	0.029	0.029	0.941	0.027	0.028
7.5	$\frac{1}{400}$	ϕ	0.5	0.506	0.053	0.053	0.474	0.064	0.069	0.507	0.053	0.054
		α_0	0.1	0.119	0.064	0.066	0.184	0.155	0.176	0.119	0.064	0.067
		α_1	0.2	0.209	0.087	0.087	0.225	0.133	0.136	0.209	0.087	0.087
		α_2	0.7	0.669	0.112	0.116	0.626	0.181	0.195	0.669	0.113	0.117
		PI	0.80	0.829	0.044	0.053	0.847	0.060	0.076	0.828	0.045	0.053
		PI	0.95	0.942	0.027	0.028	0.950	0.033	0.033	0.941	0.027	0.028
7.5	$\frac{1}{200}$	ϕ	0.5	0.506	0.053	0.053	0.447	0.068	0.086	0.507	0.053	0.054
		α_0	0.1	0.119	0.063	0.066	0.241	0.206	0.249	0.118	0.063	0.066
		α_1	0.2	0.210	0.087	0.088	0.233	0.169	0.172	0.210	0.087	0.088
		α_2	0.7	0.670	0.111	0.115	0.601	0.216	0.237	0.669	0.112	0.116
		PI	0.80	0.829	0.045	0.053	0.862	0.073	0.096	0.828	0.045	0.053
		PI	0.95	0.942	0.027	0.028	0.957	0.034	0.035	0.941	0.027	0.029
10	$\frac{1}{400}$	ϕ	0.5	0.506	0.053	0.054	0.460	0.076	0.085	0.507	0.053	0.054
		α_0	0.1	0.119	0.064	0.066	0.210	0.202	0.230	0.119	0.064	0.067
		α_1	0.2	0.211	0.087	0.088	0.260	0.201	0.209	0.211	0.087	0.088
		α_2	0.7	0.669	0.112	0.116	0.610	0.209	0.227	0.668	0.112	0.117
		PI	0.80	0.829	0.044	0.053	0.851	0.077	0.092	0.829	0.044	0.053
		PI	0.95	0.942	0.027	0.028	0.952	0.042	0.042	0.941	0.027	0.028

Table 7. Estimates for Model 2 (GARCH(1,1)-M) with T = 1000

Case				No Outlier			Outliers			Optimal Forecasts		
ψ	π		True	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
5	$\frac{1}{400}$	ϕ	0.5	0.504	0.038	0.038	0.483	0.040	0.043	0.505	0.038	0.038
		α_0	0.1	0.110	0.043	0.044	0.141	0.069	0.080	0.109	0.043	0.044
		α_1	0.2	0.203	0.058	0.058	0.200	0.068	0.068	0.202	0.058	0.058
		α_2	0.7	0.685	0.076	0.078	0.668	0.102	0.107	0.685	0.077	0.078
		PI	0.80	0.830	0.039	0.049	0.843	0.041	0.060	0.829	0.039	0.049
		PI	0.95	0.942	0.020	0.021	0.949	0.019	0.019	0.942	0.020	0.021
5	$\frac{1}{200}$	ϕ	0.5	0.504	0.038	0.038	0.465	0.041	0.054	0.505	0.038	0.038
		α_0	0.1	0.110	0.043	0.044	0.171	0.093	0.117	0.109	0.043	0.044
		α_1	0.2	0.203	0.058	0.058	0.194	0.077	0.078	0.202	0.058	0.058
		α_2	0.7	0.685	0.076	0.077	0.656	0.124	0.132	0.685	0.077	0.079
		PI	0.80	0.830	0.039	0.049	0.855	0.045	0.071	0.829	0.039	0.048
		PI	0.95	0.942	0.020	0.021	0.954	0.020	0.021	0.942	0.020	0.021
7.5	$\frac{1}{400}$	ϕ	0.5	0.504	0.038	0.038	0.463	0.046	0.059	0.505	0.038	0.038
		α_0	0.1	0.110	0.042	0.044	0.187	0.126	0.153	0.109	0.042	0.044
		α_1	0.2	0.203	0.058	0.058	0.214	0.096	0.097	0.203	0.058	0.058
		α_2	0.7	0.685	0.076	0.077	0.633	0.150	0.164	0.685	0.076	0.077
		PI	0.80	0.830	0.039	0.049	0.854	0.059	0.080	0.829	0.039	0.049
		PI	0.95	0.942	0.020	0.021	0.954	0.023	0.024	0.942	0.020	0.021
7.5	$\frac{1}{200}$	ϕ	0.5	0.504	0.038	0.038	0.431	0.049	0.085	0.505	0.038	0.038
		α_0	0.1	0.109	0.042	0.043	0.263	0.192	0.252	0.109	0.043	0.044
		α_1	0.2	0.203	0.058	0.058	0.214	0.122	0.123	0.203	0.058	0.058
		α_2	0.7	0.685	0.075	0.077	0.602	0.195	0.218	0.685	0.077	0.078
		PI	0.80	0.830	0.039	0.049	0.873	0.067	0.099	0.829	0.039	0.049
		PI	0.95	0.943	0.020	0.021	0.962	0.030	0.033	0.942	0.020	0.022
10	$\frac{1}{400}$	ϕ	0.5	0.504	0.038	0.038	0.442	0.056	0.081	0.504	0.038	0.038
		α_0	0.1	0.109	0.042	0.043	0.244	0.195	0.242	0.109	0.042	0.043
		α_1	0.2	0.204	0.058	0.059	0.243	0.144	0.151	0.204	0.058	0.058
		α_2	0.7	0.685	0.076	0.077	0.597	0.196	0.221	0.685	0.076	0.077
		PI	0.80	0.830	0.038	0.048	0.863	0.078	0.100	0.830	0.038	0.048
		PI	0.95	0.943	0.018	0.020	0.957	0.044	0.045	0.942	0.018	0.020
10	$\frac{1}{200}$	ϕ	0.5	0.503	0.038	0.038	0.399	0.059	0.117	0.504	0.038	0.038
		α_0	0.1	0.109	0.042	0.043	0.345	0.282	0.373	0.109	0.042	0.043
		α_1	0.2	0.206	0.058	0.059	0.250	0.198	0.204	0.205	0.058	0.058
		α_2	0.7	0.685	0.075	0.077	0.577	0.243	0.272	0.685	0.076	0.077
		PI	0.80	0.830	0.038	0.049	0.888	0.086	0.123	0.829	0.038	0.048
		PI	0.95	0.943	0.018	0.020	0.967	0.043	0.047	0.942	0.019	0.020

Table 8. Convergence % for the GARCH(1,1) with T = 500

ψ	5	5	7.5	7.5	10	10
π	$\frac{1}{400}$	$\frac{1}{200}$	$\frac{1}{400}$	$\frac{1}{200}$	$\frac{1}{400}$	$\frac{1}{200}$
$c\%$	99.1%	98.4%	96.6%	93.7%	92.9%	86.7%

Table 9. Convergence % for the GARCH(1,1) with T = 1000

ψ	5	5	7.5	7.5	10	10	10
π	$\frac{1}{400}$	$\frac{1}{200}$	$\frac{1}{400}$	$\frac{1}{200}$	$\frac{1}{900}$	$\frac{1}{400}$	$\frac{1}{200}$
$c\%$	99.9%	99.8%	99.0%	97.4%	98.8%	96.3%	90.8%

Table 10. Convergence % for the GARCH(1,1)-M with T = 500

ψ	5	5	7.5	7.5	10
π	$\frac{1}{400}$	$\frac{1}{200}$	$\frac{1}{400}$	$\frac{1}{200}$	$\frac{1}{400}$
$c\%$	99.5%	98.5%	96.8%	93.2%	92.5%

Table 11. Convergence % for the GARCH(1,1)-M with T = 1000

ψ	5	5	7.5	7.5	10	10
π	$\frac{1}{400}$	$\frac{1}{200}$	$\frac{1}{400}$	$\frac{1}{200}$	$\frac{1}{400}$	$\frac{1}{200}$
$c\%$	99.9%	99.8%	99.3%	97.4%	96.6%	90.6%

REFERENCES

Andersen, T. G. and Lund, J., (1997), "Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, **77**, 343-377.

Bollerslev, T., Chou, R.Y., and Kroner, K. F., (1992), "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, **52**, 5-59.

Bollerslev, T. and Mikkelsen, H. O., (1996), "Modeling and Pricing Long Memory in Stock Market Volatility," *Journal of Econometrics*, **73**, 151-184.

Bollerslev, T. and Wooldridge, J. M., (1992), "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances," *Econometric Reviews*, **11(2)**, 143-172.

Engle, R. F., (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, **50**, 987-1007.

Engle, R. F., Kane, A., and Noh, J., (1997), "Index-Option Pricing with Stochastic Volatility and the Value of Accurate Variance Forecasts," *Review of Derivatives Research*, **1**, 139-157.

Franses, P. H., and Ghijsels, H., (1999), "Additive Outliers,"GARCH and Forecasting Volatility, *International Journal of Forecasting*, **15**, 1-9.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, **48**, 1779-1802.

Hull, J. and White, A., (1987), "The Pricing of Options on Assets with Stochastic Volatility," *Journal of Finance*, **42**, 281-300.

Lee, S., and Hansen, B. E., (1994), "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, **10**, 29-52.

Lumsdaine, R. L., (1996), "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models," *Econometrica*, **64**, 575-596.

Noh, J., Engle, R. F., and Kane, A., (1994), "A Test of Efficiency for the *S&P* 500 Index Option Market using Variance Forecasts," *Journal of Derivatives*, **2**, 17-30.

Sakata, S., and White, H., (1998), "High Breakdown Point Conditional Dispersion Estimation with Application to S & P 500 Daily Returns Volatility," *Econometrica*, **66**, 529-567.

West, K. D. and Cho, D., (1995), "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics*, **69**, 367-391.