Do We Often Find ARCH Because Of Neglected Outliers?*

Philip Hans Franses
Rotterdam Institute for Business Economic Studies and
Econometric Institute, Erasmus University Rotterdam

Dick van Dijk†
Tinbergen Institute, Erasmus University Rotterdam

February 7, 1997

Abstract

In this paper we test for (Generalized) AutoRegressive Conditional Heteroskedasticity [(G)ARCH] in daily and weekly data on 22 exchange rates and 13 stock market indices using the standard Lagrange Multiplier [LM] test for GARCH and a new LM test that is resistant to additive outliers. The data span two samples of 5 years ranging from 1986 to 1995. Our main result is that we find spurious GARCH in over 50% of the cases. Using Monte Carlo simulations, in which we evaluate our empirical method, we show that this general finding indeed appears to be due to outliers. We discuss some of the implications of our findings for empirical financial modeling.

Keywords: Generalized AutoRegressive Conditional Heteroskedasticity, Lagrange Multiplier test, Outliers, Robust testing, Exchange rates, Stock market indices.

JEL classification: C12, C22, C52, G15

*We would like to thank Rob Stevense (RIBES) for his help with the data and André Lucas for some useful comments and suggestions.

†Correspondence to Dick van Dijk, Tinbergen Institute, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, `email: dvandijk@tir.few.eur.nl`

1 Introduction and motivation

An empirical regularity of many financial variables is that outliers appear in clusters, i.e., large (small) returns tend to be followed by large (small) returns of either sign. This phenomenon is observed in particular for series which are sampled daily or weekly. Since this clustering of outliers is supposed to correspond with time-varying volatility in financial indicators, there has been a growing interest in describing and forecasting volatility during the last, say, fifteen years. By far the most popular models which are used for this purpose are the so-called AutoRegressive Conditional Heteroskedasticity [ARCH] models, see Bollerslev, Chou and Kroner (1992) and Bollerslev, Engle and Nelson (1994) for recent surveys.

When standard Lagrange Multiplier [LM] tests for ARCH are used to investigate the presence of varying conditional variance in high-frequency financial time series, one almost invariably finds ARCH. de Lima (1997) argues that this result may be due to the fact that these time series seem to be characterized by heavy-tailed distributions, for which low-order moments might not exist. Using estimates of the tail index or maximal moment exponent, Jansen and de Vries (1991) and Loretan and Phillips (1993), among others, present empirical evidence that second moments of exchange rate and stock returns are finite, but that fourth order moments may not exist. de Lima (1997) shows that under these circumstances the standard LM test for ARCH rejects the null hypothesis too often, at least at conventional significance levels. To overcome this bias, Bollerslev and Wooldridge (1992) propose a modification of the LM test, which assumes distributions of the noise other than the normal distribution. Alternatively, de Lima (1997) suggests to simply trim the data before applying the standard LM test for ARCH.

In this paper we take another approach by assuming that the apparent fat tails are caused by outliers, while the underlying distribution of the error process is normal or Student t. When ARCH is to be exploited for forecasting volatility one does not want to be alarmed for ARCH when this signal is based on only a few isolated outliers. For example, it is well known to practitioners that neglecting a very short sequence of only two additive outliers [AOs] may already result in highly significant ARCH test statistics. In fact, applying ARCH models in such cases will result in systematic forecasting bias since the variance can then be largely overestimated. Because of this sensitivity to outliers, Van

Dijk, Franses and Lucas (1996) propose a robust LM test for ARCH which appears to be resistant to such AOs. Simulation results in that study show that this new LM test has good size and power properties. In the present paper we elaborate on the issue of outliers versus ARCH by comparing the results of the standard and robust LM tests for ARCH for a large data set containing 22 exchange rates and 13 stock market indices on a daily and weekly basis for two five-year long samples. Our main empirical result is that it can matter quite a lot whether one uses the standard or the robust test. To be more precise, we find that when the standard LM test points towards ARCH, the robust test suggests the same in less than 50% of the cases. Hence, it appears that the question in the title of our paper can be answered affirmatively.

The outline of this paper is the following. In Section 2, we present the two LM test statistics for ARCH. In the second part of this Section 2, we discuss the results of some Monte Carlo experiments, which give suggestions as to how we should evaluate our subsequent empirical findings. In Section 3, we briefly describe our empirical methodology, and in the second part of that section we summarize our empirical findings. In Section 4, we conclude this paper with a brief discussion of the implications for modeling volatility.

2 Testing for (G)ARCH

In this section we first briefly discuss the standard and robust LM tests for (Generalized) ARCH [GARCH], and second we illustrate their use via Monte Carlo simulations.

2.1 The LM tests

Consider the following AR(p)-GARCH(1,1) model for the return series y_t,

$$\phi(L)(y_t - \mu) = \varepsilon_t , \quad t = 1, \ldots, T, \tag{1}$$

$$\varepsilon_t = \sigma_t \eta_t, \quad \eta_t \sim i.i.d\ t_\nu \tag{2}$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{3}$$

where $\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ is a polynomial in the lag operator L, which is defined as $L^k y_t = y_{t-k}\ \forall k$, having all roots outside the unit circle. The assumption of (Student) t_ν distributed innovations η_t is quite common in applications of GARCH models to financial time series. We assume the parameters in the equation for the conditional variance of ε_t are such that this conditional variance is always positive and that the unconditional variance exists. Sufficient conditions for this are $\omega > 0$, $\alpha, \beta \geq 0$ and $\alpha + \beta < 1$. It might be argued that if the series y_t represents returns on, for example, exchange rates or stocks, the mean μ and the autoregressive parameters in $\phi(L)$ theoretically should be equal to zero to exclude arbitrage opportunities. However, when applied to series of finite length, it is customary to include an intercept in the model to allow for nonzero drift in the levels of the series during the sample period. Alternatively, the intercept sometimes is replaced by a set of dummies to capture day-of-the-week effects, see, for example, Baillie and Bollerslev (1989).

When $\beta = 0$, the model reduces to the ARCH model which was introduced by Engle (1982). Engle (1982) also derives the LM test for the null hypothesis of conditional homoskedastic errors against the alternative of ARCH(q). Lee (1991) shows that the LM test for GARCH(p,q) is exactly the same as the test for ARCH(q), when $p \leq q$. The LM test against ARCH(q) can simply be computed as TR^2, where R^2 is the coefficient of determination of an auxiliary regression of squared residuals, $\hat{\varepsilon}_t^2$ (which are obtained by estimating (1) by Ordinary Least Squares [OLS]) on an intercept and $\hat{\varepsilon}_{t-1}^2$ through $\hat{\varepsilon}_{t-q}^2$. Under the null hypothesis of no (G)ARCH, this standard LM test is asymptotically χ^2 distributed with q degrees of freedom.

Now assume that additive outliers of magnitude ζ occur with a certain probability π, such that, instead of the 'clean' series y_t, one observes the contaminated series x_t,

$$x_t = y_t + \zeta \delta_t, \tag{4}$$

where $P(\delta_t = 0) = 1 - \pi$, $P(\delta_t = 1) = P(\delta_t = -1) = \pi/2$, with $\pi \in (0, 1)$. It is well-known that OLS estimates of the mean and of the autoregressive parameters are severely biased under these circumstances. Using asymptotic arguments, Van Dijk, Franses and Lucas (1996) (hereafter DFL) show that this bias adversely affects both the size and power properties of the standard LM test for ARCH. They propose a modified LM test, which can be obtained by using a robust estimator for the parameters in (1) instead of OLS. In addition to estimates of the parameters in the conditional mean equation, the (iterative) estimation procedure suggested by DFL provides weights w_{ε_t}, which actually identify the observations which are to be considered as outliers. In fact, when $w_{\varepsilon_t} = 1$, the observation at time t is perfectly regular, while a weight smaller than one indicates that the observation does not match the properties of the bulk of the data and is somehow viewed as an outlier.

Obviously, when $w_{\varepsilon_t} = 0$, the corresponding data point is an extreme outlier. It should be noted that the 'observation at time t' here refers to the composite of the regressand y_t and regressors $y_{t-1}, \ldots, y_{t-p}$, as aberrant values of either of those can cause the weight to be smaller than 1. The modified LM test for ARCH can be computed by forming weighted regression residuals and running an auxiliary regression of the squared weighted residuals on an intercept and q lags. The LM test is then again equal to TR^2, using the R^2 from this auxiliary regression. The outlier robust LM test statistic, denoted as ξ, is computed as

$$\xi = \frac{T\hat{f}'\hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'\hat{f}}{\hat{f}'\hat{f}}, \tag{5}$$

where $\hat{Z}' = (\hat{z}_1', \ldots, \hat{z}_T')$, $\hat{z}_t = (1, w^2_{\hat{\varepsilon}_{t-1}}\hat{\varepsilon}^2_{t-1}, \ldots, w^2_{\hat{\varepsilon}_{t-q}}\hat{\varepsilon}^2_{t-q})'$, and $\hat{f}' = (\hat{f}_1', \ldots, \hat{f}_T')$, $\hat{f}_t = w^2_{\hat{\varepsilon}_t}\hat{\varepsilon}^2_t$. DFL show that this robust LM test for ARCH is again asymptotically distributed as χ^2 with q degrees of freedom. Simulation results in DFL show that the robust LM test has quite satisfactory size and power properties, already in samples as small as 100 observations, even when no outliers are present.

2.2 Some simulation results

In this section we add to the Monte Carlo evidence in DFL some new simulation experiments, which concern the frequency for which we find ARCH with the standard test, while we find no ARCH with the robust test, as well as other combinations of these individual outcomes. To investigate this, we examine two different data generating processes [DGPs]: first, a zero mean AR(1) process with homoskedastic errors [DGP I] and, second, a zero mean AR(0)-GARCH(1,1) process [DGP II]. For both DGPs, we generate 100 series of length 1250 and 1000 series of length 250. These sample sizes approximately correspond with 5 years of daily and weekly data, respectively. For each replication we record whether we find ARCH with both tests, denoted as (Y,Y), ARCH with the standard test but not with the robust test [(Y,N)], or one of the other combinations [(N,Y) or (N,N)]. These simulations will guide the interpretation of the empirical findings to be presented in the next section. In the AR(1) case, the autoregressive parameter ϕ_1 is set equal to 0.5. In the AR(0)-GARCH(1,1) case, the parameters in the conditional variance equation (3) are set equal to values which are typically found for financial time series, $\alpha = 0.15, \beta = 0.80$. The intercept in this equation, ω, is set equal to 0.05, such that the unconditional variance of ε_t is equal to the variance of η_t. In the simulations, we investigate the effects of the magnitude and frequency of occurence of AOs as in (4). We examine outliers of size $\theta = 3$, 5 and 7, which occur with probability $\pi = 0.01$, 0.05 and 0.10. In addition, the tests are computed for the uncontaminated series to obtain estimates of their size and power. The effect of the distribution of the innovations η_t is investigated as well, by setting the number of degrees of freedom of the Student t distribution equal to 5 and ∞, the latter of course corresponding to normal errors. The t_5 errors are rescaled such that they have variance equal to 1. All possible combinations of these three characteristics render eighteen different experiments per DGP and sample size. We evaluate all tests at the 5% significance level and use the asymptotic χ^2 critical values. To investigate the effect of lag length selection in the auxiliary regressions for the squared residuals used in computing the test statistics, we set q equal to 1, 5 and 10. Finally, the necessary starting values for both y_t and ε_t are set equal to zero, while the starting value for h_t in DGP II is set equal to the unconditional variance. The first 100 observations of each series are discarded to avoid dependence of our results on these starting values. Throughout, the true AR order in (1) is assumed known, while an intercept is always included in the estimation of the model under the null hypothesis.

- insert Table 1 -

The results for DGP I and $T = 250$ are reported in Table 1. From this Table, several conclusions emerge. First of all, when applied to the clean series with normal errors, the size of the robust test, which can be obtained by adding up the entries in the columns headed (Y,Y) and (N,Y), is quite satisfactory. In fact, it is even closer to the nominal 5% level than the size of the standard test, which is given by the sum of the columns (Y,Y) and (Y,N). Second, the occurence of AOs has markedly different effects on the standard and robust tests. For the standard tests, the size increases if large outliers ($\theta = 5, 7$) occur very rarely ($\pi = 0.01$). If outliers occur more frequently, the size returns to the nominal level. On the other hand, the size of the robust test is hardly affected if the probability of AOs is 1 or 5%. Only when outliers occur more frequently, the robust test deteriorates. Note that the aforementioned effects are most pronounced for $q = 1$. Finally, the results obtained when the errors are Student t_5 distributed are roughly comparable with the results for normally distributed innovations.

The comments made above are illustrated graphically in Figure 1. Plots of the estimated densities of the standard and robust LM statistics are shown, together with the appropriate asymptotic χ^2 distribution, for $q = 1$ and $\pi = 0.01$ and 0.10. From this Figure, it is seen that for $\pi = 0.01$, the distribution of the standard LM test is shifted to the right, while the same happens for the robust test when $\pi = 0.10$.

- insert Figure 1 -

Table 2 displays the Monte Carlo results for DGP I and $T = 1250$. These results are presented mainly to show that the effects of outliers on the LM tests do not disappear for larger sample sizes, but instead, these effects are amplified. This is illustrated, for example, by the fact that for this sample size both the standard and robust test tend to reject the null hypothesis too frequently for $\pi = 0.05$ as well. Also note that the size distortion in case of Student t distributed innovations becomes larger for this sample size, which seems to confirm the result in de Lima (1997).

- insert Table 2 -

The results for DGP II are shown in Tables 3 and 4, for $T = 250$ and 1250, respectively. The entries for $\pi = 0$ and $\theta = 0$ reveal that the main disadvantage of the robust test is a considerable drop in power when no outliers are present for the smaller of the two sample sizes considered. The difference with the power of the standard test is approximately 20% for $T = 250$. This illustrates that protection against aberrant observations comes at a cost. If outliers do occur, the situation is completely reversed however. Except for very rare and small AOs, the power of the standard test decreases dramatically, while the power of the robust test remains the same or even becomes higher.

- insert Tables 3 and 4 -

Summarizing, the main conclusion which emerges from our Monte Carlo experiments is that if the time series are contaminated with AOs, the inference from evaluating both the standard and the robust LM tests for ARCH shows two typical patterns. First, in case of homoskedastic errors, the combination (Y,N) is the prevalent outcome, i.e., the standard test rejects the null hypothesis, while the robust test does not (see Tables 1 and 2). Second, in case GARCH is present while there are also outliers, both tests again tend to find the opposite, although now (N,Y) becomes the dominant empirical outcome. These typical outcomes will be compared with our empirical results for actual financial data in the next section.

3 ARCH in exchange rate and stock market returns

In this section we consider the joint application of the two LM tests for ARCH for 35 financial time series. In Section 3.1 we discuss the data and our empirical research methodology, and in Section 3.2 we report on our empirical findings.

3.1 Data and methodology

We consider 35 financial time series, which are sampled either daily or weekly, for a ten-year period ranging from 1986 to 1995. We compute our tests for two subsamples, 1986-1990 and 1991-1995, each of which contain 5 years of data. In terms of daily and weekly data, this means that we have samples of about 1250 and 250 observations, respectively. Given this choice to split the sample, we have one sample containing the 1987 stock market crash. Results for other (overlapping) samples, which turn out to be qualitatively similar, are not reported here to save space, but are available from the authors. The 35 series can be grouped into 22 exchange rates (versus the US dollar) and 13 stock market indices. The exchange rates concern the Austrian shilling, Australian dollar, Belgian franc (commercial), British pound, Canadian dollar, Danish kroner, ECU, Finnish markka, French franc, Greek drachme, Irish pound, Japanese yen, Malaysian ringgit, Dutch guilder, New Zealand dollar, Norwegian kroner, South African rand (commercial), Singapore dollar, Spanish peseta, Swedish kroner, Swiss franc and the German Dmark. The stock markets concern those in Brussels (BSE), Amsterdam (CBS), Frankfurt (DAX), New York (Dow Jones), London (FTSE), Hong Kong (Hang Seng), Tokyo (Nikkei), Madrid (MSE), Milan (MC), USA (S&P500), Singapore, Taipei, and Stockholm (VEC). We apply the LM tests for ARCH while setting q equal to 1, 5 and 10. The tests are applied to both the raw and 'prewhitened' series. In both cases the series are demeaned ('demedianed') first by subtracting the mean (median) before applying the standard (robust) test to the series. For the daily series, daily means and medians are used in order to allow for possible day-of-the-week effects. The series are prewhitened by fitting an AR model of order 5 and 2 to the daily and weekly series, respectively. We also compute the tests when the AR order in (1) is selected by the Akaike and Schwarz Information Criteria. This does not yield qualitatively different results, and hence we do not report them. All tests are evaluated at the 5% significance level. Similar results are obtained using 1% and 10% significance levels, which therefore are not displayed. We summarize our empirical findings by recording the number of times both tests find ARCH [(Y,Y)], the standard test finds ARCH while the robust test does not [(Y,N)] and vice versa [(N,Y)], and how often both tests do not find ARCH [(N,N)].

3.2 Results

Table 5 presents some general results for the raw and prewhitened daily and weekly series.

- insert Table 5 -

For the daily series, interpreting the outcomes is not straightforward. On the one hand, depending on the number of lagged squared residuals included in the auxiliary regressions, both tests reject the null hypothesis for 29-57% of the series. Given the results from the simulations presented in the previous section, this definitely points towards the presence of ARCH in these series. On the other hand however, the robust test does not find ARCH while the standard test does in about 38-70% of the cases, as shown by the entries in the last column of this table. This suggests that ARCH effects can often be caused by the occurrence of outliers.

For the weekly series, both tests are unable to reject the null for 47-62% of the series, providing rather strong evidence for the absence of ARCH at this sampling frequency. For the majority of the remaining series, the standard LM test finds ARCH, while the robust test does not. Again, this corresponds with our above Monte Carlo results, and suggests that apparent ARCH effects in these weekly data may very well be the consequence of only a few relatively large AOs.

- insert Table 6 -

In Table 6, the results for the tests applied to the raw series are displayed at a more disaggregated level, by focusing on the two subsamples and on exchange rates and stock market returns individually. It appears that the evidence for ARCH is more convincing in the 1986-1990 sample, as the entries in the (Y,Y) column in general are larger than for the subsample comprising 1991-1995. Also note that for the first subsample the number of stock indices for which the robust test does not find ARCH while the standard test does is remarkably high. This suggests that the standard test perhaps is 'fooled' by exceptional events, like the stock market crash of October 1987.

To illustrate that exceptional outliers are important, consider the final two columns of this table, which contain percentages of observations which receive weights w_{ε_t} equal to zero and equal to one in the robust estimation procedure which is used to compute the robust LM statistic. It can be seen from the last column that between 3.5% and 8.1% of the observations are downweighted, i.e., obtain a weight smaller than one, of which between 2.0% and 5.5% are discarded completely (which is what effectively is done when an observation receives weight zero). Note that this does not imply that the data contain these fractions of outliers, because the robust estimation procedure always downweights some observations, even if all data points are perfectly regular. The only conclusion we can draw based on these percentages is that the robust estimation procedure considers only a relatively small number of observations to be dramatically aberrant, and also that the spurious finding of ARCH can be caused by only a small number of very large outliers.

4 Conclusions

In this paper we have explored the possibility that ARCH effects, which are commonly found in high-frequency financial time series, are caused by only a few aberrant observations. By comparing the outcomes of standard and robust LM tests for ARCH to a large number of empirical time series, we conclude that often this indeed is a likely possibility. We therefore recommend to apply both the standard and robust tests when considering the application of GARCH models, as the joint outcome of these tests might be rather informative with regard to the appropriateness of this class of models. In case only the standard test finds ARCH, the possibility of AOs instead of ARCH has to be given some serious thought. Only if both the standard and robust tests point towards ARCH, it is warranted to estimate a GARCH model without preliminary removal of outliers. If only the robust test rejects the null hypothesis, ARCH effects can be hidden by outliers and then it might be worthwhile to use a robust estimation method for ARCH models.

Our conclusion has serious implications for financial modeling, especially in the areas of option valuation and risk management, among others. For example, the assumption of constant volatility underlying the Black-Scholes and other option valuation models has been under attack in recent years. In this paper we showed that this might not be a completely unreasonable assumption after all, since constant volatility may have been rejected because of a few neglected outliers.

References

Baillie, R.T. and T. Bollerslev (1989) The message in daily exchange rates: a conditional variance tale, *Journal of Business and Economic Statistics*, **7**, 297–305.

Bollerslev, T. and J.M. Wooldridge (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews*, **11**, 143–172.

Bollerslev, T., R.F. Engle and D.B. Nelson (1994) ARCH models, in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics IV*, 2961–3038, Elsevier Science, Amsterdam.

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.

de Lima, P.J.F. (1997) On the robustness of nonlinearity tests to moment condition failure, *Journal of Econometrics*, forthcoming.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.

Jansen, D. and C. de Vries (1991) On the frequency of large stock returns: putting booms and busts into perspective, *Review of Economics and Statistics*, **73**, 18–24.

Lee, J.H.H. (1991) A Lagrange multiplier test for GARCH models, *Economics Letters*, **37**, 265–271.

Loretan, M. and P.C.B. Phillips (1993) Testing the covariance stationarity of heavy-tailed time series: an overview of the theory with applications to several financial datasets, *Journal of Empirical Finance*, **1**, 211–248.

van Dijk, D., P.H. Franses and A. Lucas (1996) Testing for ARCH in the presence of additive outliers, Econometric Institute Report 9659/A, Erasmus University Rotterdam.

Table 1: Simulation evidence of ARCH, using standard and robust LM tests: the case of no ARCH[1]

			$\nu=\infty$				$\nu=5$			
π	θ	q	(Y,Y)	(Y,N)	(N,Y)	(N,N)	(Y,Y)	(Y,N)	(N,Y)	(N,N)
.0	0	1	1.4	1.5	2.8	94.3	0.4	3.0	4.9	91.7
		5	1.7	1.4	3.4	93.5	0.4	5.1	4.3	90.2
		10	2.4	1.6	2.8	93.2	0.4	5.5	4.3	89.8
.01	3	1	1.9	9.9	4.1	84.1	0.5	7.2	5.3	87.0
		5	1.6	7.9	3.1	87.4	0.2	7.1	5.0	87.7
		10	1.1	6.2	4.1	88.6	0.3	7.2	5.1	87.4
	5	1	1.9	29.7	3.4	65.0	0.9	18.9	4.3	75.9
		5	1.2	15.5	4.2	79.1	0.7	15.3	3.3	80.7
		10	0.9	13.2	4.8	81.1	0.8	12.2	3.8	83.2
	7	1	1.9	29.6	3.0	65.5	1.1	23.5	4.0	71.4
		5	0.9	15.3	4.7	79.1	0.8	15.8	3.2	80.2
		10	0.7	13.6	4.6	81.1	1.0	14.1	3.6	81.3
.05	3	1	2.9	11.3	8.7	77.1	1.3	8.7	9.8	80.2
		5	0.7	7.9	8.3	83.1	0.9	7.0	7.5	84.6
		10	0.7	6.6	5.6	87.1	0.5	6.6	7.3	85.6
	5	1	0.9	9.1	8.1	81.9	0.4	8.1	5.9	85.6
		5	0.3	7.2	6.4	86.1	0.4	6.5	5.7	87.4
		10	0.5	7.7	5.5	86.3	0.4	7.4	4.6	87.6
	7	1	0.5	5.6	5.0	88.9	0.3	4.7	5.6	89.4
		5	0.1	5.9	4.7	89.3	0.3	6.6	4.4	88.7
		10	0.2	7.5	4.8	87.5	0.5	6.7	4.2	88.6
.10	3	1	2.5	6.1	11.6	79.8	1.3	6.3	13.5	78.9
		5	0.9	5.3	7.8	86.0	0.5	5.6	9.4	84.5
		10	0.6	5.7	7.4	86.3	0.5	6.0	7.8	85.7
	5	1	1.8	2.8	33.1	62.3	0.6	5.3	15.9	78.2
		5	1.0	4.6	19.4	75.0	0.6	4.1	11.2	84.1
		10	0.6	4.4	15.4	79.6	0.5	4.2	8.3	87.0
	7	1	0.2	3.2	8.4	88.2	0.3	4.4	9.5	85.8
		5	0.0	4.0	6.4	89.6	0.3	3.9	6.1	89.7
		10	0.2	4.9	4.4	90.5	0.1	3.7	5.6	90.6

[1] Evidence of ARCH using the standard LM test and using the new LM test, which is robust to additive outliers. The cells report the frequency of a certain outcome when the test statistics are evaluated at the 5% significance level. For example, (Y,N) means that the standard LM test detects ARCH (Y) while the robust test does not (N). The series are generated by an AR(1) process with autoregressive parameter $\phi = 0.5$ and t_ν distributed errors. The q is the number of lags included in the auxiliary regressions for the squared residuals from an AR(1) regression. The table is based on 1000 replications for sample size $T = 250$, which roughly corresponds with 5 years of weekly data.

Table 2: Simulation evidence of ARCH using standard and robust LM tests: the case of no ARCH[1]

			$\nu = \infty$				$\nu = 5$			
π	θ	q	(Y,Y)	(Y,N)	(N,Y)	(N,N)	(Y,Y)	(Y,N)	(N,Y)	(N,N)
.0	0	1	1.0	5.0	3.0	91.0	0.0	6.0	7.0	87.0
		5	2.0	1.0	11.0	86.0	1.0	5.0	4.0	90.0
		10	1.0	2.0	4.0	93.0	0.0	9.0	7.0	84.0
.01	3	1	5.0	32.0	7.0	56.0	5.0	12.0	5.0	78.0
		5	4.0	23.0	7.0	66.0	1.0	12.0	7.0	80.0
		10	2.0	15.0	3.0	80.0	1.0	16.0	6.0	77.0
	5	1	10.0	84.0	1.0	5.0	6.0	71.0	2.0	21.0
		5	9.0	60.0	3.0	28.0	3.0	47.0	3.0	47.0
		10	2.0	56.0	3.0	39.0	2.0	38.0	5.0	55.0
	7	1	8.0	89.0	1.0	2.0	8.0	83.0	0.0	9.0
		5	9.0	62.0	2.0	27.0	4.0	68.0	2.0	26.0
		10	2.0	59.0	2.0	37.0	2.0	49.0	5.0	44.0
.05	3	1	27.0	25.0	24.0	24.0	21.0	20.0	22.0	37.0
		5	9.0	24.0	24.0	43.0	7.0	17.0	24.0	52.0
		10	3.0	20.0	19.0	58.0	8.0	12.0	19.0	61.0
	5	1	10.0	31.0	15.0	44.0	9.0	28.0	12.0	51.0
		5	4.0	21.0	14.0	61.0	1.0	28.0	6.0	65.0
		10	1.0	14.0	11.0	74.0	2.0	16.0	8.0	74.0
	7	1	2.0	18.0	17.0	63.0	5.0	12.0	12.0	71.0
		5	1.0	11.0	12.0	76.0	2.0	10.0	11.0	77.0
		10	1.0	9.0	11.0	79.0	1.0	7.0	11.0	81.0
.10	3	1	20.0	6.0	45.0	29.0	21.0	10.0	53.0	16.0
		5	6.0	8.0	37.0	49.0	6.0	9.0	40.0	45.0
		10	4.0	3.0	25.0	68.0	5.0	11.0	30.0	54.0
	5	1	13.0	2.0	75.0	10.0	8.0	4.0	33.0	55.0
		5	7.0	3.0	69.0	21.0	2.0	6.0	19.0	73.0
		10	1.0	5.0	65.0	29.0	0.0	7.0	18.0	75.0
	7	1	4.0	1.0	37.0	58.0	3.0	2.0	24.0	71.0
		5	0.0	5.0	24.0	71.0	1.0	5.0	12.0	82.0
		10	1.0	2.0	21.0	76.0	0.0	6.0	11.0	83.0

[1] Evidence of ARCH using the standard LM test and using the new LM test, which is robust to additive outliers. The cells report the frequency of a certain outcome when the test statistics are evaluated at the 5% significance level. For example, (Y,N) means that the standard LM test detects ARCH (Y) while the robust test does not (N). The series are generated by an AR(1) process with autoregressive parameter $\phi = 0.5$ and t_ν distributed errors. The q is the number of lags included in the auxiliary regressions for the residuals from an AR(1) regression. The table is based on 100 replications for sample size $T = 1250$, which roughly corresponds with 5 years of daily data.

Table 3: Simulation evidence of ARCH using standard and robust tests: the case of ARCH[1]

			$\nu = \infty$				$\nu = 5$			
π	θ	q	(Y,Y)	(Y,N)	(N,Y)	(N,N)	(Y,Y)	(Y,N)	(N,Y)	(N,N)
.0	0	1	37.6	27.7	6.4	28.3	12.7	35.2	9.1	43.0
		5	58.9	23.5	1.6	16.0	21.9	45.2	6.8	26.1
		10	57.0	22.8	2.6	17.6	20.2	46.1	8.2	25.5
.01	3	1	25.8	21.5	15.5	37.2	10.0	25.6	11.9	52.5
		5	43.1	17.6	15.1	24.2	16.5	32.8	13.2	37.5
		10	38.5	19.4	17.1	25.0	15.1	34.9	12.2	37.8
	5	1	12.9	9.5	30.3	47.3	5.8	15.7	17.2	61.3
		5	22.5	7.7	38.3	31.5	10.2	21.0	19.2	49.6
		10	21.6	8.6	38.0	31.8	9.9	21.8	18.0	50.3
	7	1	6.6	4.7	36.4	52.3	3.7	10.1	19.2	67.0
		5	13.6	4.8	47.0	34.6	6.4	13.7	23.0	56.9
		10	14.1	5.9	45.4	34.6	7.2	15.8	20.6	56.4
.05	3	1	9.5	13.4	21.1	56.0	6.3	15.8	17.5	60.4
		5	16.3	12.5	29.3	41.9	11.5	19.2	20.5	48.8
		10	15.0	14.3	27.6	43.1	9.9	20.5	20.3	49.3
	5	1	3.8	4.0	40.1	52.1	2.7	8.4	22.1	66.8
		5	7.5	3.0	57.1	32.4	6.7	8.0	28.8	56.5
		10	8.7	3.1	53.4	34.8	6.7	9.3	27.5	56.5
	7	1	1.8	2.6	41.5	54.1	1.9	5.6	22.4	70.1
		5	4.6	2.0	58.8	34.6	3.9	6.1	30.5	59.5
		10	5.5	2.5	56.1	35.9	3.9	6.7	29.9	59.5
.10	3	1	3.0	11.1	18.1	67.8	3.3	12.9	17.4	66.4
		5	6.0	12.9	22.6	58.5	6.0	15.4	22.8	55.8
		10	6.2	11.4	21.1	61.3	5.9	16.2	21.7	56.2
	5	1	3.1	2.5	43.6	50.8	1.9	5.3	26.2	66.6
		5	5.4	1.9	60.1	32.6	4.5	3.9	34.1	57.5
		10	4.9	1.7	58.0	35.4	4.5	5.1	32.6	57.8
	7	1	2.0	1.5	43.8	52.7	1.5	3.7	25.3	69.5
		5	3.3	1.5	61.2	34.0	3.1	3.5	34.4	59.0
		10	2.7	1.4	60.0	35.9	3.1	3.6	34.2	59.1

[1] Evidence of ARCH using the standard LM test and using the new LM test, which is robust to additive outliers. The cells report the frequency of a certain outcome when the test statistics are evaluated at the 5% significance level. For example, (Y,N) means that the standard LM test detects ARCH (Y) while the robust test does not (N). The series are generated by a GARCH(1,1) process with parameters $\alpha = 0.15$, $\beta = 0.8$ and t_ν distributed errors. The q is the number of lags included in the auxiliary regressions used in computing the LM statistics. The table is based on 1000 replications for sample size $T = 250$, which roughly corresponds with 5 years of weekly data.

Table 4: Simulation evidence of ARCH using standard and robust tests: the case of ARCH[1]

			$\nu=\infty$				$\nu=5$			
π	θ	q	(Y,Y)	(Y,N)	(N,Y)	(N,N)	(Y,Y)	(Y,N)	(N,Y)	(N,N)
.0	0	1	97.0	3.0	0.0	0.0	73.0	23.0	3.0	1.0
		5	100.0	0.0	0.0	0.0	94.0	6.0	0.0	0.0
		10	99.0	1.0	0.0	0.0	94.0	5.0	1.0	0.0
.01	3	1	97.0	2.0	1.0	0.0	70.0	25.0	4.0	1.0
		5	100.0	0.0	0.0	0.0	93.0	6.0	1.0	0.0
		10	100.0	0.0	0.0	0.0	95.0	4.0	1.0	0.0
	5	1	58.0	1.0	40.0	1.0	51.0	19.0	20.0	10.0
		5	71.0	0.0	29.0	0.0	77.0	6.0	17.0	0.0
		10	70.0	0.0	30.0	0.0	79.0	4.0	17.0	0.0
	7	1	18.0	0.0	80.0	2.0	29.0	9.0	43.0	19.0
		5	32.0	0.0	68.0	0.0	47.0	1.0	47.0	5.0
		10	30.0	0.0	70.0	0.0	54.0	3.0	42.0	1.0
.05	3	1	68.0	7.0	21.0	4.0	57.0	19.0	18.0	6.0
		5	86.0	1.0	13.0	0.0	82.0	6.0	12.0	0.0
		10	85.0	0.0	15.0	0.0	84.0	5.0	10.0	1.0
	5	1	14.0	0.0	83.0	3.0	24.0	11.0	49.0	16.0
		5	20.0	0.0	80.0	0.0	48.0	1.0	50.0	1.0
		10	24.0	0.0	76.0	0.0	47.0	0.0	53.0	0.0
	7	1	9.0	0.0	88.0	3.0	11.0	4.0	62.0	23.0
		5	9.0	0.0	91.0	0.0	29.0	0.0	69.0	2.0
		10	10.0	0.0	90.0	0.0	24.0	0.0	75.0	1.0
.10	3	1	40.0	12.0	27.0	21.0	35.0	19.0	39.0	7.0
		5	60.0	6.0	31.0	3.0	69.0	5.0	23.0	3.0
		10	55.0	4.0	37.0	4.0	65.0	6.0	26.0	3.0
	5	1	8.0	0.0	92.0	0.0	17.0	8.0	60.0	15.0
		5	12.0	0.0	88.0	0.0	35.0	2.0	60.0	3.0
		10	14.0	0.0	86.0	0.0	36.0	2.0	59.0	3.0
	7	1	5.0	0.0	95.0	0.0	11.0	0.0	66.0	23.0
		5	5.0	0.0	95.0	0.0	23.0	1.0	72.0	4.0
		10	4.0	0.0	96.0	0.0	22.0	0.0	74.0	4.0

[1] Evidence of ARCH using the standard LM test and using the new LM test, which is robust to additive outliers. The cells report the frequency of a certain outcome when the test statistics are evaluated at the 5% significance level. For example, (Y,N) means that the standard LM test detects ARCH (Y) while the robust test does not (N). The series are generated by a GARCH(1,1) process with parameters $\alpha = 0.15$, $\beta = 0.8$ and t_ν distributed errors. The q is the number of lags included in the auxiliary regressions used in computing the LM statistics. The table is based on 100 replications for sample size $T = 1250$, which roughly corresponds with 5 years of daily data.

Table 5: Evidence of ARCH in exchange rate and stock market data, some overall results[1]

	q	(Y,Y)	(Y,N)	(N,Y)	(N,N)	(Y,N)/((Y,Y)+(Y,N))
Raw series						
Daily data	1	18	41	2	9	0.70
	5	36	27	3	4	0.43
	10	39	24	4	3	0.38
Weekly data	1	8	15	3	44	0.65
	5	13	21	3	33	0.62
	10	11	16	6	37	0.59
Prewhitened series						
Daily data	1	26	32	3	9	0.55
	5	40	24	2	4	0.38
	10	43	21	3	3	0.33
Weekly data	1	8	16	3	43	0.67
	5	10	22	6	32	0.69
	10	11	16	7	36	0.59

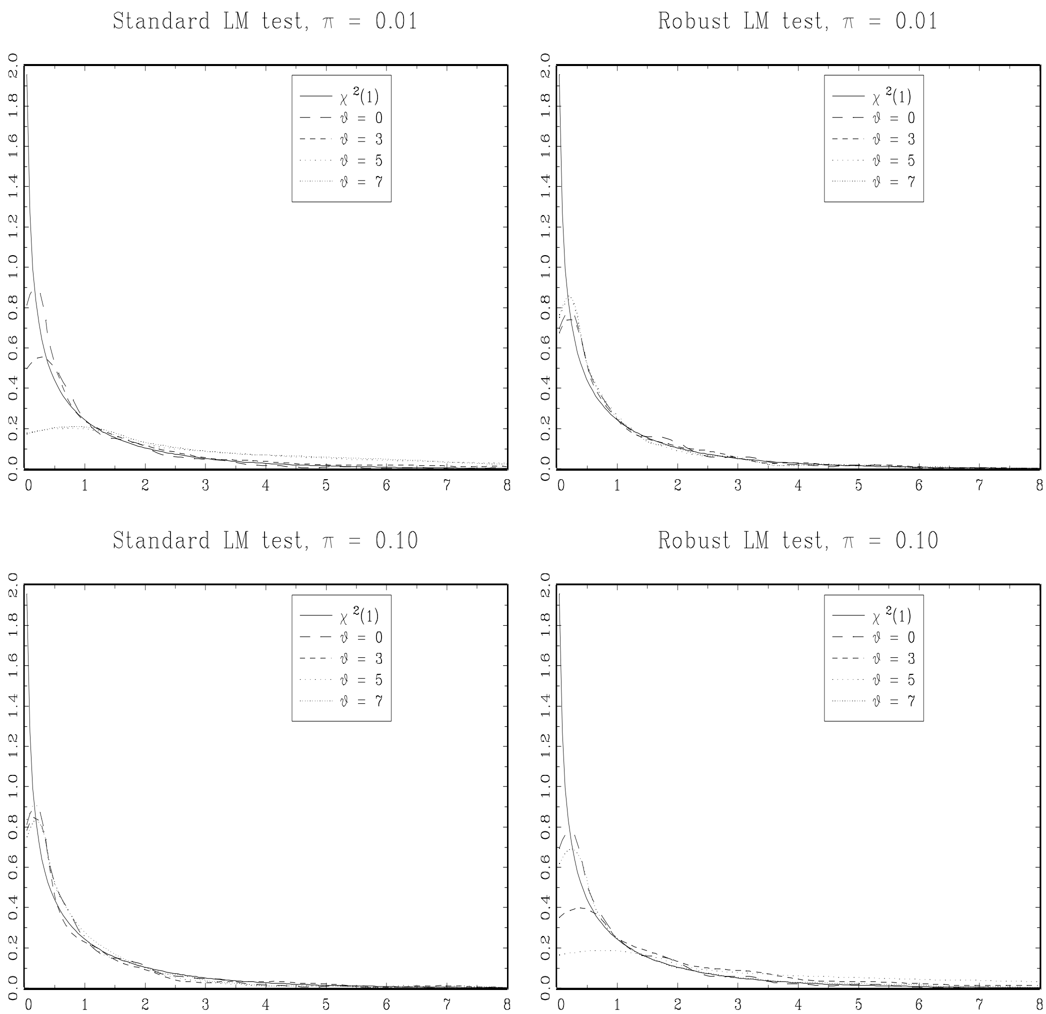
[1] Evidence of ARCH using the standard LM test and using the new LM test, which is robust to additive outliers. The cells report the number of times (out of 70 cases: 2 samples of 22 exchange rates and 13 stock indices) a certain outcome appears when the test statistics are evaluated at the 5% significance level. For example, (Y,N) means that the standard LM test detects ARCH (Y) while the robust test does not (N). (Y,N)/((Y,Y)+(Y,N)) denotes the frequency that the robust ARCH test does not find ARCH while the standard ARCH test does. The data and the empirical methodology are presented in Section 3.1.

Table 6: Evidence of ARCH in exchange rate and stock market data[1]

Sample	q	(Y,Y)	(Y,N)	(N,Y)	(N,N)	Weight 0	Weight 1
Exchange rates, daily data							
1986-1990	1	4	13	0	5	4.5	91.9
	5	10	10	1	1		
	10	12	10	0	0		
1991-1995	1	2	18	0	2	3.3	94.0
	5	8	11	1	2		
	10	9	10	2	1		
Exchange rates, weekly data							
1986-1990	1	3	5	3	11	2.2	96.4
	5	5	6	1	10		
	10	4	3	5	10		
1991-1995	1	0	1	0	21	4.2	94.1
	5	0	5	1	16		
	10	0	4	0	18		
Stock markets, daily data							
1986-1990	1	9	3	1	0	4.2	93.3
	5	10	3	0	0		
	10	10	2	1	0		
1991-1995	1	3	7	1	2	2.5	95.2
	5	8	3	1	1		
	10	8	2	1	2		
Stock markets, weekly data							
1986-1990	1	5	8	0	0	5.2	92.6
	5	4	8	1	0		
	10	5	5	2	1		
1991-1995	1	0	2	0	11	2.0	96.5
	5	1	3	3	6		
	10	2	4	0	7		

[1] Evidence of ARCH using the standard LM test and using the new LM test, which is robust to additive outliers. The cells in columns 3-6 report the number of times (out of 22 cases for the exchange rates and 13 cases of the stock market indices) a certain outcome appears when the test statistics are evaluated at the 5% significance level. For example, (Y,N) means that the standard LM test detects ARCH (Y) while the robust test does not (N). The final two columns give the percentage of observations which receive weights smaller than 0.05 and larger than 0.95 in the robust estimation procedure. The data and the empirical methodology are presented in Section 3.1.

Figure 1: Empirical Distributions of LM tests for ARCH



Note: Kernel estimates of the empirical distributions of the standard and robust LM tests for ARCH(1). The figures are based on 1000 replications of DGP I, for sample size $T = 250$. The Epanechnikov kernel with automatically selected bandwidth is used.