Detecting Local Outliers in Financial Time Series

Peter Verhoeven[a] and **Michael McAleer[b]**

[a]School of Economics and Finance, Curtin University of Technology, WA 6102, Australia
(VerhoevP@cbs.curtin.edu.au)
[b]Department of Economics, University of Western Australia, WA 6009, Australia

Abstract: In this paper the concept of local outliers is introduced to volatility modeling. It is demonstrated that local outliers are influential observations which have a substantially greater impact on the QMLE of the GARCH model than other observations. Local outliers are not detected by the Chen and Liu (1993) method as they are relatively small in size and do not generate large residuals for the GARCH model. When a simple filter for these outliers is applied to the data, substantial gains in efficiency are obtained.

Keywords: *Local outliers, global outliers, conditional volatility, influential observations, filtering, outlier-ness.*

1. INTRODUCTION

1.1 Local Outliers

Whether to treat an observation as an outlier has been, and remains, an important problem in statistical inference. There is no universally-accepted definition of an outlier. Hawkins (1980) proposed the following intuitive definition: "An outlier is an observation that deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism." According to this *model-based* definition, outliers are a "subjective, post-data concept" (Beckman and Cook, 1983) which can only exist with reference to a specific model. For autoregressive models, clustering and outlier detection are closely related concepts.

Breunig et al. (1999) introduced the notion of *local* outliers by taking into account the clustering structure in its bounded neighborhood, specifically the k nearest (local) neighbours. The concept of local outliers is similar to the concept of *spatial* outliers, which are observations that are inconsistent with (or isolated from) those in their immediate neighbourhood, even though they may be consistent with the overall sample (Shekhar et al., 2001). This contrasts with the concept of *global* outliers, namely observations that are isolated from the overall sample.[1] The concept of local outliers may be considered as a measure of continuity, namely how well the next observation can be predicted beyond the known (previous) observations.[2]

Breunig et al. (1999) also suggest that being an outlier is not just a binary property, and introduce the concept of a local outlier *factor* to capture the relative degree of isolation or *outlier-ness* of an observation with respect to its surrounding neighbours. For many situations, it is more meaningful to assign to each observation a degree of being an outlier as this allows observations to be ranked according to this degree, thereby providing an order for analyzing outliers.

1.2 Specification of Mean and Variance

Consider the following generalised autoregressive conditional heteroskedasticity model, GARCH(1,1), where the conditional mean (or logarithmic-returns) is given by an AR(1) process:

$$y_t = \mu + \varphi y_{t-1} + \varepsilon_t, \quad |\varphi| < 1 \qquad (1)$$

where $\varepsilon_t \sim f\left(0, \sqrt{h_t}\right)$, and the conditional variance of the residuals is given by a GARCH(1,1) process:

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} \qquad (2)$$

The conditional variance of ε_t can be used to obtain the normalized (or standardized) error, $\eta_t = \varepsilon_t / \sqrt{h_t}$, which is assumed to be i.i.d. $f(0,1,\theta)$. The α (or ARCH) parameter captures the short run persistence of shocks, while the β (or GARCH) parameter captures the contribution of shocks to long run persistence, α+β. Sufficient conditions for the

[1] Thus, global outliers refer to observations that lie far outside the tails of the *unconditional* distribution.

[2] Continuity assumes that there is a greater chance of connection between two contiguous elements (auto correlated) than between either one of the elements and any of the other non-contiguous elements.

conditional variance to be positive are $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$.

In these models, kurtosis of the conditional distribution has been widely used as a diagnostic check for the correct specification of the mean-variance dynamics. A given observation will appear to be extreme in terms of the standardized residuals η_t, when its magnitude is large *relative* to the conditional volatility, $\sqrt{h_t}$. Conversely, an extreme observation will lie in the tails of the conditional distribution when the conditional volatility at time t is high, that is, when ε_{t-1}^2 or $\sqrt{h_{t-1}}$ is large.

It follows that the *influence* of an observation at time t on the Quasi Maximum Likelihood Estimates (QMLE) largely depends on whether it is clustered with its *preceding* neighbours, that is, whether or not it can be anticipated. In the absence of clustering, the QMLE will try to capture isolated observations by increasing h_t through increasing the α or β estimates.[3]

Whether an observation persists, that is, clustered with its nearest *proceeding* neighbours, may also be an important factor in determining its influence on the QMLE. When an observation is isolated from its preceding neighbours but is not persistent, an increase in either (or both) the α or β estimates may lead to overestimation of the subsequent volatility, leading to inliers[4] in the conditional distribution. That is, it may result in a peaked conditional distribution, thereby yielding high kurtosis.[5]

In summary, the degree of influence of an observation at time t on the QMLE of the conditional variance model will depend on: (i) lagged conditional volatility; (ii) lagged squared residual; and (iii) whether the observation persists, that is, whether ε_{t+i}^2 is large. For example, the smaller is the difference in squared magnitude between the observation and its preceding neighbours, the smaller is the marginal increase in the α estimate required to capture it through $\alpha\varepsilon_{t-1}^2$. Similarly, the larger is the previous conditional volatility (h_{t-1}), the smaller is the marginal increase in the β estimate required to capture such an observation through βh_{t-1}. However, there will be a trade-off between the α and β estimates so as not to cause over-persistence in the conditional variance, as implied by the magnitude of the α+β estimate. More specifically, for non-persistent observations, an increase in the α estimate may be more desirable than an increase in the β estimate, as this is more likely to generate normality in the standardised residuals.

This paper proposes a simple intuitive measure for the degree of outlier-ness of an observation based on its degree of clustering. The remainder of the paper is organized as follows: Section 2 presents the methodology and data, Section 3 discusses the empirical results, and Section 4 concludes the paper.

2. METHODOLOGY AND DATA

In order to obtain a better understanding of what makes an observation influential, a single observation of various sizes will be inserted into the various estimation periods. Two alternative methodologies will be used, with one involving a fixed estimation period and the other using a rolling sample.

In the fixed estimation period, a single global outlier will be inserted at various (random) points.[6] The aim of this method is to examine the degree of outlier-ness and influence of an observation on the QMLE by taking into account the characteristics of the data in its immediate surroundings. The fixed estimation period consists of a single period of 2500 daily returns, allowing a comparison between the original and adjusted data. A rolling window methodology of 1240 observations is used to examine how various features of the data further from the added extreme observation mitigates its degree of influence.

The following measures are proposed for the degree of outlier-ness of an observation:

$$cva = \frac{\varepsilon_t^2}{\frac{1}{k}\sum_{i=1}^{i=k}\varepsilon_{t+i}^2} \tag{3}$$

$$cvb = \frac{\varepsilon_t^2}{\frac{1}{k}\sum_{i=1}^{i=k}\varepsilon_{t-i}^2} \tag{4}$$

$$vratio = \frac{cva}{cvb} = \frac{\sum_{i=1}^{i=k}\varepsilon_{t-i}^2}{\sum_{i=1}^{i=k}\varepsilon_{t+i}^2} \tag{5}$$

[3] The α and β estimates are obtained subject to the restriction that α+β< 1.
[4] Inliers are observations that have values very close to the mean of the distribution.
[5] Strictly speaking, when such inliers appear at a frequency that cannot reasonably be considered to be consistent with normality, these should be classified as outliers.

[6] The global outlier has a magnitude of 10 standard deviations from the unconditional sample mean.

The outlier factors cvb and cva measure the degree of isolation of the variance of the observation from its nearest preceding and proceeding neighbours, respectively. Hence, the larger are these factors, the more isolated will the observation be from its immediate neighbourhood. Vratio is defined as the ratio of the conditional volatility (or clustering) preceding and proceeding the observation. The smaller is this ratio, the greater is the persistence of the observation. In this paper, the length of the neighbourhood (k) is set at 10 days.

In order to establish the relevance of these outlier factors in determining the influence of observations on the QMLE of the GARCH model, simple regressions are performed where the QMLE are regressed on these factors. The QMLE is obtained under the assumption of conditional normality.

The empirical analysis is based on the daily logarithmic returns of the Dow Jones Industrial Average (DJIA) and the National Association of Securities Dealers Automated Quotation (Nasdaq) Composite Index. The sample consists of 2500 observations for the period 24 August 1990 to 28 July 2000.

3. EMPIRICAL RESULTS

3.1 Single Additive Outlier

The influence of an observation on the QMLE of the mean-variance model depends on its degree of outlier-ness from the preceding neighbourhood, that is, whether or not it can be scaled adequately by h_t. This outcome depends on: (i) whether conditional volatility immediately preceding the extreme observation is sufficiently large[7]; and (ii) whether it is preceded by a sufficiently large observation (ε_{t-1}^2). When either of these conditions is met, its degree of outlier-ness (cvb) is expected to be small. Thus, a slight increase in either the α or β estimate is likely to lead to a sufficient increase in h_t so as to capture the observation. A mitigating factor will be whether or not the observation persists, otherwise the effect of inliers generated as a result of an increase in h_{t+i} will outweigh the effect of outliers. For example, if the observation has a high degree of outlier-ness with respect to its preceding observations and its volatility does not persist, the α and β estimates may be biased downwards so as to reduce the number of inliers.[8]

The results confirm our intuition, namely that the influence of a global outlier on the QMLE of the GARCH model can be substantial, though the extent of its influence depends largely on its degree of outlier-ness. In most instances, when a global outlier is inserted, the α estimate increases substantially in magnitude. For Nasdaq (DJIA), of 20 (23) cases where a single global outlier (10σ) is inserted, there are: (i) 14 (11) cases where there is an increase in the α estimates; (ii) 5 (9) cases where there is an increase in the β estimates; (iii) 13 (13) cases where there is a decrease in the ($\alpha+\beta$) estimates; (iv) 4 (8) cases where there is an increase in both the β and ($\alpha+\beta$) estimates; and (v) 3 (3) cases where there is an increase in both the α and ($\alpha+\beta$) estimates. Interestingly, there are no cases in which both the α and β estimates increased, and just 1 case (3 cases) where both the α and β estimates decreased. These results imply that, although the penalty on outliers is substantially larger than the penalty on inliers, there is a strong trade-off between the α and β estimates in controlling the number of inliers that are generated.

The relationship between excess kurtosis (Ke) and the outlier factor (cvb) for Nasdaq is approximately S-shaped and is steepest when the outlier factor is between 100 and 1000. Outside this region, the relationship is relatively flat. Hence, when an observation has an outlier factor that exceeds 100, it becomes increasingly difficult to capture, as demonstrated by the exponential increase in excess kurtosis.[9]

The results confirm the finding that a major determinant of the influence of an observation on the QMLE is its degree of isolation with respect to its preceding neighbours (cvb) rather than its absolute size. For example, when a global outlier is inserted at position 2435, it has no significant influence on either the QMLE or the excess kurtosis of the standardised residuals, implying that it can be adequately captured by the model. When the same observation is inserted at position 1805, it causes a substantial increase in the α estimates and in excess kurtosis. The only difference between these two observations is the degree of clustering with their immediate neighbourhoods. That is, the global outlier inserted at position 2435 has a substantially smaller outlier factor (cvb = 4.95) than the global outlier inserted at position 1805 (cvb = 119.1). These results also show that the influence of an observation on the QMLE

[7] This is the case when an extreme observation appears several steps beforehand, thereby causing masking effects.

[8] This is because an increase in either the α or β estimate will not be sufficient to capture the extreme observation.

[9] As expected, there is a very strong negative relationship between the maximized log-likelihood (MLL) values and the excess kurtosis (Ke) of the standardised residuals: the greater is the excess kurtosis, the stronger is the violation of normality, the poorer its fit, and the lower are the MLL values.

increases exponentially with the logarithm of the outlier factor. In general, observations with an outlier factor less than 10 have very little influence on the QMLE and on the excess kurtosis.

Generally, there is a very strong negative univariate relationship between: (i) α and β estimates; (ii) $(\alpha+\beta)$ estimates and excess kurtosis; and (iii) $(\alpha+\beta)$ estimates and cvb. There is no obvious relationship between the α or β estimates and either cvb or excess kurtosis.

The results of the multiple regressions show that, for both series, excess kurtosis (Ke) is strongly negatively related to the α, β, and $(\alpha+\beta)$ estimates and to vratio, and positively related to cva. Furthermore, the $(\alpha+\beta)$ estimates are negatively related to both the α estimates and cva, and positively related to both the β estimates and vratio. This implies that the less persistent is the outlier (as measured by vratio), the smaller are the $\alpha+\beta$ estimates. Collectively, the results suggest that: (i) global outliers are captured by increasing the α or β estimates; and (ii) more persistent outliers are more easily accommodated.

Although the Nasdaq series are substantially more volatile (40%) than the DJIA series, the regression results are qualitatively quite similar. For DJIA, the α estimates are negatively related to cvb, while the β estimates are negatively related to cva. However, when both outlier factors are included in the regression equation, the α and β estimates become negatively related to cva. This suggests that the less persistent is the outlier, the smaller are the α and β estimates, which is consistent with the results for the $(\alpha+\beta)$ estimates.

Presumably, the cost in generating multiple inliers (at subsequent positions) is larger than the benefit of capturing a single outlier. When $(\alpha+\beta)$ is included in the regression equations for both α and β, α is marginally significant and negatively related to cvb, whereas β is marginally significant and positively related to cvb. None of the GARCH parameters is now significantly related to cva. These results may be explained by the fact that cva and $(\alpha+\beta)$ are significantly negatively correlated with each other (-0.469). Both the α and β estimates are strongly positively related to vratio, irrespective of whether the β and α estimates are included in the regression equations, respectively. However, after controlling for the effects of $(\alpha+\beta)$, α becomes significantly negatively related to vratio, while the reverse is true for β. Hence, there are substantial trade-offs between the α and β estimates, which are primarily driven by the persistence of the outlier.

For both series, ω is positively related to both cva and the α estimates, and negatively related to both the $(\alpha+\beta)$ estimates and vratio. This suggests that the QMLE may capture non-persistent outliers by increasing the conditional volatility through increasing the ω estimates.

In conclusion, the results confirm our intuition that the degree of outlier-ness of an observation is a strong determinant of its degree of influence on the QMLE of the conditional volatility model. Moreover, there appears to be a strong trade-off between the α and β estimates, and between outliers and inliers. In particular, the estimator attempts to capture local outliers by increasing the α or β estimates, depending on the degree of persistence. When the observation is isolated from its preceding and proceeding neighbours (with large cvb and cva values), the estimator reduces the number of inliers by decreasing the $(\alpha+\beta)$ estimates through reducing both the α and β estimates.

3.2 Volatility of the Series

The impacts of a global outlier on the QMLE are also expected to be determined by the features of the data that are further from the observation. There are two main reasons for this result, namely: (i) all observations contribute to the GARCH estimates and determine whether an outlier can be captured through h_t; and (ii) the impacts of a global outlier on the QMLE will affect the volatility forecasts of all other observations.

In order to examine the impacts of the features of the remaining data on the influence of a global outlier, the rolling window methodology is employed. For Nasdaq, a single global outlier (10σ) is inserted at a position (1245) where it has a high local outlier factor (cvb = 351.2, cva = 390.22).

The results show that the QMLE of the conditional variance model are affected by the inserted outlier as soon as it appears in the estimation period. In particular, there is: (i) a sharp decrease in the α estimate (from 0.119 to 0.065); (ii) a sharp increase in the β estimate (from 0.726 to 0.938); (iii) a sharp increase in the $(\alpha+\beta)$ estimate (from 0.845 to 1.003); and (iv) a sharp increase in excess kurtosis (from 5.307 to 22.855). When the outlier is eventually purged from the estimation period: (i) the α estimate abruptly decreases (from 0.186 to 0.162); (ii) the β estimate increases (from 0.794 to 0.823); (iii) the

($\alpha+\beta$) estimate increases (from 0.979 to 0.985); and (iv) excess kurtosis decreases (from 6.337 to 4.062). These results suggest that the influence of the inserted outlier on the GARCH estimates remains substantial, even when it appears at the start of the estimation period.

For the original data, as the window slides forward into a period of higher conditional volatility, the α, β and ($\alpha+\beta$) estimates increase substantially in size. The result is a lower excess kurtosis for the standardised residuals. Generally, as more volatile data enter the estimation period, the influence of the inserted outlier on the β and ($\alpha+\beta$) estimates decreases substantially, whereas its influence on the α estimate remains large.

In conclusion, the influence of a local outlier on the GARCH estimates depends not only on the conditional volatility of its local surroundings, but also on the conditional volatility of the other observations. In particular, the more volatile are the data, the larger are the QMLE of the parameters of the conditional variance model, the better can the outlier be captured by the model, and the less influential are outliers.

3.3 Filtering Local Outliers

Without any formal method for detecting the position of random jumps in the conditional variance, we proceed with an informal method of outlier filtering. In this section, we examine whether down-weighting those observations that have a local outlier factor (cvb) greater than some arbitrary set critical value (c) can lead to improvements in estimation.

We proceed by down-weighting those observations that have a local outlier factor that exceed some preset threshold level.[10] This yields the *uncontaminated* series. The larger is the threshold level, the greater is the probability that the observation is generated by a process that is different from GARCH. The following method of adjustment is used:

$$\varepsilon_t^* = 3\sqrt{\frac{1}{k}\sum_{i=1}^{i=k}(\varepsilon_{t-i}-\bar{\varepsilon})^2}$$

For purposes of comparison, we also apply the Chen and Liu (1993) filter to the residuals of the GARCH(1,1) model. The ten largest local outliers are listed in Table 1.

[10] The threshold values used are 5, 10, 20 and 30.

Table 1. The ten largest local outliers in volatility

Nasdaq

Cvb	y_t*		η_t
63.568	-0.04337	(-3.284)	-6.268
38.079	-0.02604	(-1.972)	-4.488
34.702	-0.07274	(-5.509)	-5.859
31.228	-0.03710	(-2.810)	-4.840
31.073	-0.05714	(-4.327)	-4.835
29.470	-0.02937	(-2.224)	-4.322
22.050	-0.01485	(-1.125)	-2.251
19.890	-0.01907	(-1.444)	-2.541
19.800	0.02467	(1.868)	2.611
19.696	-0.08954	(-6.780)	-4.166

DJIA

Cvb	y_t*		η_t
119.130	-0.04006	(-4.311)	-6.650
34.030	-0.07455	(-8.022)	-6.698
32.530	-0.02456	(-2.643)	-4.698
31.630	0.02958	(3.183)	3.641
27.160	0.01274	(1.371)	1.953
24.750	-0.02016	(-2.169)	-3.080
24.670	-0.01549	(-1.667)	-2.603
22.930	-0.01727	(-1.858)	-2.288
22.290	-0.01871	(-2.013)	-3.636
20.800	0.04466	(4.806)	4.913

*The size of the observations in terms of their distance in standard deviations from the mean is given in parentheses.

Interestingly, a significant majority of global outliers[11] does not appear as outliers in the standardised residuals. For example, for the volatile Nasdaq series, only 7 of a total of 39 global outliers appear as outliers in standardized residuals.[12] For Nasdaq, only 6 of a total of 44 global outliers have a local outlier factor greater than 10, whereas 19 of a total of 20 outliers in the conditional distribution have a local outlier factor greater than 10.[13] This implies that the local outlier factor, as measured by cvb, is an accurate indicator of an observation's outlier-ness in terms of its likelihood of becoming an outlier in the standardised residuals.

It is also worth noting that the majority of observations that can be considered as outliers in the standardised residuals (that is, have a local outlier factor larger than 10) have a negative sign. For example, for Nasdaq, 19 of 20 outliers (95%) in the standardised residuals are negative, while for DJIA,

[11] For convenience, in this section we consider observations that are more than 3 standard deviations removed from the sample mean global outliers.

[12] For the DJIA series, 15 of a total of 30 global outliers in the unconditional returns are classified as global outliers when standardised by their conditional volatility.

[13] For DJIA, 12 of a total of 30 global outliers have a local outlier factor greater than 10, whereas 24 of a total of 26 outliers in the conditional distribution have a local outlier factor greater than 10.

20 of 26 outliers (77%) in the standardised residuals are negative. Similarly, for both series, 70% of the observations that have a local outlier factor greater than 10 are negative. In contrast, the proportion of global outliers that is negatively or positively signed is approximately equal.

It is important to note that local outliers can rarely be classified as belonging to the class of the largest events. For example, for Nasdaq (DJIA), only 12% (29%) of the observations with a local outlier factor greater than 10 are larger than 3σ, with more than two-thirds (one-third) being smaller than 2σ. Most of the observations with a local outlier factor greater than 10 belong to the least volatile periods in the series (1990-1995), implying that bad news, rather than good news, is more likely to arrive unexpectedly and during quiet periods.

The local outlier factor seems to be an effective way of identifying those observations that are likely to contribute significantly to excess kurtosis and have a substantial influence on the QMLE. As an example, for Nasdaq (DJIA), 38% (50%) of the observations with a local outlier factor greater than 10 are outliers in the standardised residuals. For Nasdaq (DJIA), the 6 (3) observations that have the largest η_t values correspond to the 6 (3) observations with the largest local outlier factors. Overall, only 4 of 46 observations with the largest η_t values have a local outlier factor less than 10.

Although not reported in detail here, the results for the QMLE of GARCH(1,1) and GARCH-t for Nasdaq for the original and adjusted data show that the residual kurtosis can be reduced substantially by down-weighting just a few (7) local outliers. Furthermore, when a large number of observations with a high outlier factor are down-weighted, the net impact on the α and β estimates is only marginal. In contrast, the Chen and Liu (1993) filter leads to substantially decreased α estimates and increased β estimates. This finding is consistent with the fact that the Chen and Liu filter identifies global outliers which, by definition, are very large in magnitude. Since such global outliers are frequently clustered with extreme observations, their local outlier factor is typically small. As the marginal influence of local outliers on the QMLE's of the conditional variance can be of either sign, the net impact of filtering a large number of these outliers on the QMLE of GARCH is generally quite small.

A major benefit of filtering local outliers is that there are substantial gains in efficiency, as indicated by the markedly larger t-ratios for both the α and βestimates. This contrasts with the results of the Chen and Liu (1993) filter. The findings of the outlier-robust GARCH-t model are closer to that of the Chen and Liu (1993) filter than that of the local outlier filter.

As expected, down-weighting local outliers has no substantial impacts on the forecast errors. This is consistent with the findings that: (i) local outliers are frequently small observations; and (ii) down-weighting a large number of local outliers has no significant or consistent influence on the GARCH estimates.

4. CONCLUSION

In this paper we proposed the notion of a local outlier and the degree of outlier-ness for volatility modelling. It was found that the most obvious (global) outliers are not necessarily the most influential observations as they are often clustered, indicating ARCH effects rather than isolated outliers. In particular, we showed that local outliers rather than global outliers have significant impacts on the GARCH estimates. Moreover, it was shown that identifying and filtering local outliers may lead to a substantial improvement in the efficiency of the GARCH estimates.

5. ACKNOWLEDGEMENT

The second author wishes to acknowledge the financial support of the Australian Research Council.

6. REFERENCES

Beckman, R.J. and R.D. Cook (1983), "Outlier...s", Technometrics, 25, 119-149.

Breunig, M.M., H.P. Kriegel, R.T. Ng, and J. Sander (1999), "OPTICS-OF: Identifying local outliers", in Lecture Notes in Computer Science, Volume 1704, 262-270 (1999), Springer Verlag, Berlin.

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.

Hawkins, D.M. (1980), "Identification of outliers", Chapman and Hall, London.

Shekhar, S., Y. Huang, W. Wu, C.T. Lu, and S. Chawla (2001), "What's spatial about spatial data mining: three case studies", in Data Mining for Scientific and Engineering Applications, R.L. Grossman, C. Kamath, P. Kegelmeyer, V. Kumar and R.R. Namburu (eds.), 487-514, Kluwer, New York.