Forecasting Commodity Market Volatility in the Presence of Extreme Observations

C. Watkins and M. McAleer

Department of Economics, University of Western Australia (cwatkins@ecel.uwa.edu.au)

Abstract Extreme observations in commodity returns time series data occur as the result of shocks to a market through macroeconomic news, or market-specific events including fundamental and speculative pressures. However, outliers can have a dominating and deleterious effect in empirical models. This paper examines the forecasting of returns volatility in the presence of extreme observations using an AR(1)-GARCH(1,1) model for a non-ferrous metal futures contract. A simple method of accommodating extreme observations is applied that involves squeezing outliers to various thresholds. The forecasts obtained using this method are compared with a simple model in which all observations from the sample are used, and no adjustment for atypical observations is made. Estimates from the rolling one-step ahead models are presented graphically, and a number of forecast evaluation criteria are used to compare the forecasts generated under different outlier regimes.

Keywords: Extreme observations; Outliers; GARCH; Futures contracts; Volatility; Squeezing data

1. VOLATILITY, OUTLIERS AND METALS

Volatility in commodity markets represents risk to producers and consumers of commodities. Risk in storable commodity markets is manifest as uncertainty for producers in terms of revenues, for consumers in terms of costs, and for stock holders in terms of margins. Derivatives, such as futures and options, are routinely used to hedge against price risk in commodity markets. Strategies for hedging, and pricing of options and other derivatives, require knowledge of the volatility of the underlying time series.

As volatility is unobservable, it must be estimated. The modelling and forecasting of volatility has received much attention in the literature since the development of the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle [1982], as well as Bollerslev's [1986] inclusion of a lagged dependent variable to form the Generalised ARCH (GARCH) model. A multitude of GARCH-type models has since been developed to incorporate various stylised facts of volatility, and the time series properties of financial returns. Empirical research has followed the development of these theoretical models, but GARCH(1,1) remains the most widely used time-varying volatility model in practice.

Although GARCH is adequate in forecasting standard volatility, it is apparent that the model does not predict outlying observations particularly well. Furthermore, in the presence of extreme observations, the model performs poorly in terms of parameter estimation and forecasts. Under these circumstances, forecasts of volatility will be affected by outliers, and could possibly be improved by adjusting the outlying and extreme observations. It may be argued, however, that extreme observations contain useful information about possibly important, and frequently non-repetitive, events. Such aberrant observations can be associated with asymmetric returns, spikes and clusters in volatility, and hence should not be deleted from the sample altogether.

As for many commodity and financial returns series, returns on industrially-used non-ferrous metals futures contracts contain extreme observations. Metals prices embody a business cycle component. It is to be expected, therefore, that macroeconomic news will effect daily returns on metals futures contracts and trading on the spot market, as do market-specific events including fundamental and speculative pressures. Hedge fund activities are a common source of large price fluctuations for commodity spot and futures prices, particularly for aluminium.

This paper examines the forecasting of volatility in returns in the presence of extreme observations using an AR(1)-GARCH(1,1) model for a non-ferrous metal (specifically, aluminium) futures contract traded on the London Metal Exchange

(LME). A method of accommodating extreme observations is used, namely simply squeezing the extreme returns observations lying beyond some threshold back to the threshold. Chen and Liu [1993] suggested an iterative weighting algorithm for adjusting extreme observations for ARMA models, which was applied to GARCH models in Franses and Ghijsels [1999]. The advantage of the method used in this paper over that originally suggested by Chen and Liu [1993] is its straightforward and mathematically simple nature. The forecasts obtained using the squeezing method are compared with a model in which all observations from the sample are used, and no adjustment for atypical observations is made. Estimates from rolling one-step-ahead forecasting models are presented graphically, and a number of forecast evaluation criteria are used to compare the forecasts generated under different outlier regimes. Improved forecasting of volatility in metals markets will allow superior risk management by producers and consumers of industrial metals, improved hedging strategies and options pricing, and the provision of enhanced information for speculators and investment funds.

2. ACCOMMODATING EXTREME OBSERVATIONS

The GARCH model of Bollerslev [1986] is used in this paper, specifically the AR(1)-GARCH(1,1) model. The conditional mean of futures price returns in this model is given by:

$$r_t = \mu + \varphi r_{t-1} + \varepsilon_t, \quad |\varphi| < 1 \tag{1}$$

and the conditional variance of ε_t is:

$$\varepsilon_t = \eta_t \sqrt{h_t} \tag{2}$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \tag{3}$$

where r_t denotes returns on futures price from period t-1 to t; ε_t is the unconditional shock; η_t is a sequence of normally, independently and identically distributed random variables with zero mean and unit variance; and h_t is the conditional variance of returns. For the GARCH process to exist, the conditional variance must be positive, so that $\omega > 0$, $\alpha \geq 0$, and $\beta \geq 0$.

Several statistical properties have been established for the GARCH(p,q) process in order to define the moments of the unconditional shock. Ling and McAleer [2001] derived a necessary and sufficient condition for the existence of the moments of a family of GARCH processes described in He and Teräsvirta [1999], which includes the GARCH(1,1) model. Moment conditions for the GARCH(1,1) process can easily be checked.

The necessary and sufficient condition for the second moment to exist for the GARCH(1,1) model, guaranteeing the GARCH(1,1) process is strictly stationary and ergodic, is given by:

$$\alpha + \beta < 1. \tag{4}$$

The fourth moment of the unconditional shock will exist if and only if the following condition is satisfied (assuming normality):

$$3\alpha^2 + 2\alpha\beta + \beta^2 < 1. \tag{5}$$

In section 4, the existence of second and fourth moment conditions is examined under various outlier regimes.

GARCH modelling in the presence of outlying observations has been examined in a number of recent papers, importantly, Hotta and Tsay [1998], Franses and Ghijsels [1999], Franses and van Dijk [1999], Yew et al. [2001], and Verhoeven and McAleer [2000]. An outlier is typically defined as an observation from a different population to that of the sample, while an extreme observation is an atypical observation originating from the same population as the rest of the sample. In an empirical sense, the distinction between an outlier and an extreme observation is to some extent arbitrary. Several types of outliers may occur in financial time series, including additive outliers, innovation outliers, level shifts and variance changes. In the case of an additive outlier, only one observation is affected. Several observations are affected where there is an innovation outlier. The specific definition of an outlying observation depends on the model of outliers used.

Outliers can have a dominating and deleterious effect on the (quasi-) maximum likelihood estimates of GARCH parameters, leading to model misspecification, poor forecasts and invalid inferences. Biases may be induced upwards for the estimate of the ARCH parameter and downwards for the estimate of the GARCH parameter. The tendency of GARCH to over-predict the persistence of moderate to high observations is increased when outliers are present in the estimation window. Adverse effects on the size and power of the LM test used to detect ARCH effects are likely, leading to an increased likelihood of model misspecification. A small number of isolated and/or clustered outliers may result in spurious ARCH effects when none is present. Rolling forecasts may suffer temporary detrimental effects due to carry-over effects of outliers in previous estimation windows, and to permanent effects of model misspecification and parameter estimates that are biased for the entire sample. Temporary effects are particularly noticeable when extreme observations enter and leave the estimation window.

This paper uses a computationally straightforward method to accommodate extreme observations proposed in Yew et al. [2000], continuing a theme developed by Franses and Ghijsels [1999]. Daily returns observations lying beyond a threshold are deemed to be extreme, and are squeezed back to the threshold itself prior to estimating the model. The threshold is defined as a multiple of the sample standard deviation on either side of the sample mean. In this application, several alternative multiples of the standard deviation are used as thresholds to transform the data. For comparison purposes, estimates and forecasts are also generated from a model for which the data are not transformed. The outlier regimes and associated thresholds are:

Regime A: No transformation;
Regime B: 4 standard deviation threshold;
Regime C: 3 standard deviation threshold;
Regime D: 2.5 standard deviation threshold.

As a rolling AR(1)-GARCH(1,1) model is used in all cases, the sample mean and sample standard deviation are calculated for each rolling window. Each sample window is transformed individually, prior to estimating the model and generating forecasts from that window.

3. NON-FERROUS METALS DATA

The LME is the major international market for the main industrially-used non-ferrous metals, namely aluminium, aluminium alloy, copper, lead, nickel, tin, and zinc[1]. The Exchange began trading aluminium futures in 1978, but the market was initially thin. At that stage, the aluminium market was dominated by a small number of large producers. Internationally, price was set on a producer list basis. These list prices changed infrequently, but transactions prices were often substantially discounted relative to the list price. Over the 1980's, the LME contract gained an increasing volume of trade as the industry moved to pricing on the basis of market quotations. Such quotations from the LME are now used by the aluminium industry worldwide as the basis for pricing. Industry participates in the futures market on the LME extensively for hedging. Commodity and hedge funds are active in the aluminium futures market, and have been blamed for periods of high volatility due to speculative activities. The aluminium market is now the largest metals market by value and contract turnover on the LME.

Daily data for 3-month contract settlement prices are obtained for aluminium that cover the period 1 October 1982 to 31 May 2000, providing a total of 4458 observations. Prices quoted by the LME prior to July 1993 are denominated in British Pounds. The 3-month futures prices are converted from British Pounds to US Dollars using the 3-month US Dollar to British Pound exchange rate. After July 1993, prices are quoted by the exchange in US Dollars. The returns series are calculated as:

$$r_{t-1,t} = \left(f_t - f_{t-1}\right)/f_{t-1}. \tag{6}$$

A plot of the price and returns series is presented in Figure 1. Several extreme observations are apparent, the largest of which is an extreme negative observation on 19 October 1987, followed by a positive correction on the following day. The correction is followed by another large negative observation on 29 October and a positive correction on 30 October, with yet another large negative observation on 2 November. A substantial cluster of volatility extending for approximately two years follows the largest negative outlier. Within this cluster, there are several large negative and positive outliers. Several clusters of volatility are apparent at other points in the data set, and there are numerous extreme observations, the majority of which are negative. The late 1990's is notable in that any extreme observations are small relative to those which occurred in the 1980's and early 1990's. The price series show a number of apparent structural breaks. Descriptive statistics for the returns show the series are leptokurtic, but not skewed.

4. APPLICATION TO VOLATILITY FORECASTING

Inspection of the data and testing for stylised facts are important first steps to determine which model best describes the conditional variance of aluminium returns. The AR(1)-GARCH(1,1) model is appropriate for symmetric and leptokurtic returns, and adequately represents the data in terms of rolling diagnostic tests for normality, serial correlation, and the existence of ARCH effects. Estimates of the conditional variance were not sensitive to the choice of model for the mean equation. Furthermore, choice of the GARCH(1,1) model allows an evaluation of outlier adjustment regimes for the most commonly used volatility model in empirical research.

The AR(1)-GARCH(1,1) model is estimated for each returns series using a rolling window of 500 observations, which rolls 3958 times over a sample of 4458 observations. The procedure is programmed using EViews, and generates coefficient estimates, standard errors, t-ratios, moment conditions and one-step ahead out-of-sample forecasts. Estimates from the rolling samples are treated as "data" in the descriptive discussion below, where each model is estimated by maximum likelihood. The forecast volatility is compared with the 'true' volatility calculated over the same window, where the 'true' volatility is defined as:

$$v_t = \left(r_t - \bar{r}\right)^2 \tag{7}$$

where v_t refers to the 'true' volatility at time t, and

[1] Futures contracts for the precious metal silver have also been traded on the LME since late 1999.

$\bar{r}$ is the mean of returns over the window for the sample used. The 1-day ahead forecast error, u_{t+1} is defined as:

$$u_{t+1} = v_{t+1} - \hat{h}_{t+1}. \qquad (8)$$

Forecasts of volatility generated by the models are compared using the following criteria, namely, mean error (ME), mean absolute error (MAE), root mean squared error (RMSE), smoothed mean absolute percentage error (SMAPE), smoothed weighted median absolute percentage error (SMedWAPE), and smoothed weighted mean absolute percentage error (SWMAPE).

4.1 Parameter Estimates and Moment Conditions

Plots of the rolling α (or ARCH) coefficient estimates are provided in Figures 2 (regime A), 3 (regime B), 4 (regime C) and 5 (regime D), and rolling β (or GARCH) estimates are shown in Figures 6 (regime A), 7 (regime B), 8 (regime C), and 9 (regime D)[2]. All parameter estimates are positive, as required for the GARCH(1,1) model. As expected, most ARCH parameter estimates are small at around 0.1, while most GARCH parameter estimates are large at around 0.9. However, there are numerous instances where outlying observations have a clear and detrimental influence on the parameter estimates. It is apparent in the plots of coefficient estimates derived from regime A that outliers have an upward impact on the ARCH estimates and a downward impact on the GARCH estimates. Over the entire sample, the estimates show substantial variability when no adjustment for outliers is undertaken. In general, as the threshold is increased, from regime B to C to D, the ARCH and GARCH parameter estimates become less variable.

Estimate number 811 is generated when the October 1987 outliers, discussed in section 3, have only recently entered the estimation window. Figures 2 and 6 show that these outliers have a substantial effect on the ARCH and GARCH parameter estimates, respectively, under the no outlier adjustment regime (regime A). The ARCH plots show an upwards spike, while the GARCH plots show a downward spike. The absolute magnitude of the movement in each estimate from their previous levels is similar, so that the second moment is hardly affected. Squeezing the October 1987 outliers to 4 standard deviations from the mean (regime B in Figures 3 and 7) reduces the spike in both the ARCH and GARCH plots to a level where it is almost unnoticeable. Further squeezing of the outliers has little effect, as can be seen from the plots for regimes C (Figures 4 and 8) and D (Figures 5 and 9).

The same holds for nearby spikes apparent in the ARCH and GARCH plots. For example, estimate 921 is substantially different from its neighbouring ARCH and GARCH estimates. The October 1987 outliers remain within the estimation window at estimate 1274, where there is a small spike in both coefficient estimates. In both cases, regime B shows a substantial decrease in the difference between the affected estimate and its neighbours. Regimes C and D make little difference beyond that shown in the plots for regime B.

Between estimates 1784 and 2295, the ARCH estimate shifts upward dramatically from around 0.1, which is typical for ARCH, to 0.5, which is highly unusual. Over the same period, the GARCH estimates fall to a minimum of under 0.1. This episode coincides with 17 October 1991, when the largest positive outlier enters the estimation window. Interestingly, this returns observation is not followed by a negative correction. Some 65 trading days later, there is a second large positive return, which is also not followed by a correction. When the October 1991 outlier leaves the window, the ARCH estimates return to reasonable levels, as do the GARCH estimates. As the threshold for adjusting the data is tightened, regimes B, C and D show that this outlier has a successively lower impact on the ARCH and GARCH estimates. Squeezing the returns data to a threshold of 2.5 standard deviations (regime D) has the greatest impact on decreasing the ARCH estimates and on increasing the GARCH estimates. However, while the largest positive outlier remains within the estimation window, the ARCH estimates are still far higher than is normally expected, and the GARCH estimates are far lower.

In the late 1990's (with observation 3346 representing 2 January 1996), where there are few extreme observations in the returns series, the GARCH estimates become extremely variable, and plots of the t-ratios show the estimates are not significant. For much of this period, the ARCH estimates are relatively stable, but they are not significant. There is little difference between the four outlier adjustment regimes over these estimates in terms of parameter stability.

Applying the outlier adjustment regimes to the returns data also reduces the variability observed in the t-ratios associated with the rolling ARCH and GARCH estimates. Spikes in the series of t-ratios, apparently induced by outlying observations, are reduced in magnitude. The t-ratios become more stable over the rolling windows, but in almost all cases this does not change the significance of the estimates at the 5% level.

[2] For reasons of space, plots of t-ratios for the ARCH and GARCH estimates, and second and fourth moments, are not presented.

Table 1. Forecast Evaluation.

Evaluation Criteria	2.5 SD	3 SD	4 SD	No Adjustment
ME	0.002465	0.002780	0.003155	0.003528
MAE	0.006815	0.006938	0.007117	0.007309
MedAE	0.005884	0.006032	0.006189	0.006299
RMSE	0.009134	0.009239	0.009411	0.009666
RMSE(-)	0.011973	0.011862	0.011896	0.011661
RMSE(+)	0.007720	0.008015	0.008366	0.008886
RMedSE	0.005884	0.006032	0.006189	0.006299
SMAPE	74.01	74.57	75.35	75.98
SMedAPE	62.54	63.29	64.07	64.93
SMWAPE	52.94	52.39	52.06	51.82
SMedWAPE	30.78	31.09	31.81	32.24
% Forecasts Under	28.45	27.61	25.97	25.37
% Forecasts Over	71.55	72.39	74.03	74.63
% 2MC Satisfied	96.13	95.28	93.66	92.47
% 4MC Satisfied	92.57	92.04	91.51	89.84

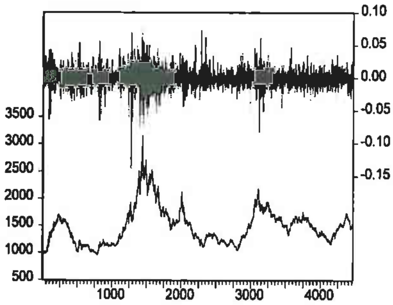

Figure 1. Returns and Prices.

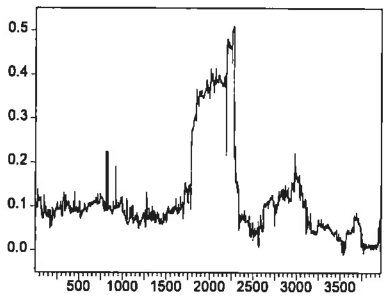

Figure 2. α Estimates-Regime A.

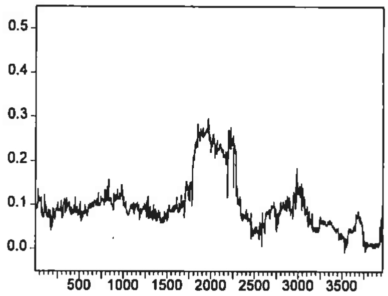

Figure 3. α Estimates-Regime B.

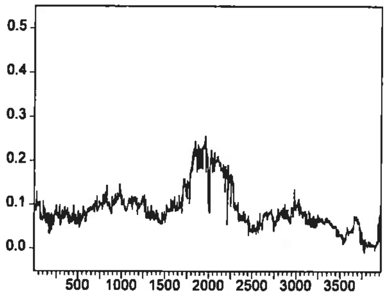

Figure 4. α Estimates-Regime C.

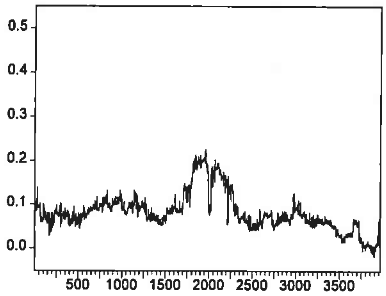

Figure 5. α Estimates-Regime D.

Figure 6. β Estimates-Regime A.

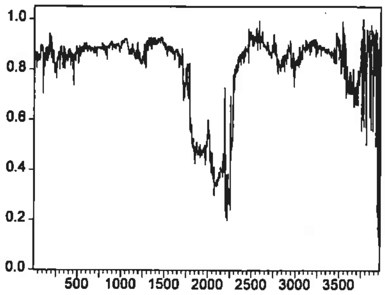

Figure 7. β Estimates-Regime B.

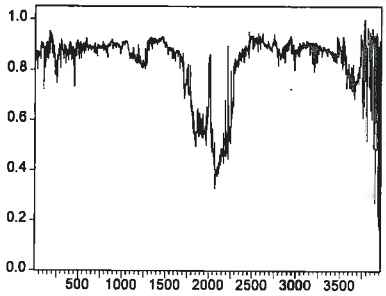

Figure 8. β Estimates-Regime C.

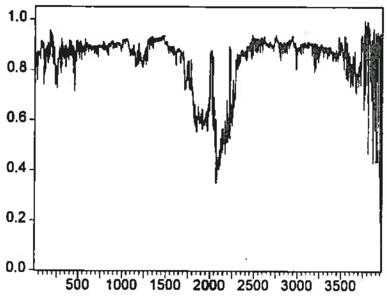

Figure 9. β Estimates-Regime D.

The percentages of estimation windows for which the second and fourth moment conditions are satisfied under various outlier regimes are shown in Table 1. Accommodation of outlying observations improves slightly the percentage of windows that satisfy each regularity condition. For the second moment, the proportion of windows satisfying the condition increases from around 92.5% under regime A to over 96% under regime D. For the fourth moment, the improvement is slightly smaller, from just under 90% for regime A to over 92.5% for regime D. The tighter is the threshold, the higher is the number of windows satisfying the conditions.

4.2 Forecasting Performance

Table 1 shows the forecasting performance according to several forecast evaluation criteria for the AR(1)-GARCH(1,1) model under each outlier adjustment regime. ME, MAE, MedAE and RMSE all provide some evidence to support an improvement in the forecast performance of the GARCH(1,1) model when outlying observations are squeezed to their respective thresholds. Importantly, the forecast performance improves as the thresholds are narrowed, so that regime D performs the best, followed by regimes C, B, and A. The model using the raw data with no adjustment for outliers (regime A) performed the worst in terms of forecasts.

RMSE was also applied to only positive forecast errors, denoted RMSE(+), and to only negative forecast errors for RMSE(-). While adjusting extreme observations improved the GARCH(1,1) model with respect to over-forecasting, the model generated larger under-forecasting errors. With no outlier adjustment, the forecasting model over-predicts volatility almost 75% of the time, and under-predicts around 25%. Over the 3958 estimation windows, the outlier adjustment regimes decrease the number of over-forecasts and increase the number of under-forecasts. The tighter is the threshold, the lower is the likelihood of the GARCH(1,1) model to over-forecast.

Several smoothed forecast evaluation measures, where the forecast error is normalised on the mean of the forecast and actual volatilities, were also calculated. SMAPE and SMedAPE indicated a modest improvement in forecast performance by accommodating extreme observations, and this improved with the tightness of the threshold. Weighting smoothed measures by the size of the actual volatility relative to its mean produced conflicting results. While SMedWAPE showed an improvement due to accommodating outliers, SMWAPE indicated that forecasting performance deteriorated upon using the outlier adjustment regimes. For periods of high volatility, squeezing outlying observations under regimes B, C or D is detrimental to the forecasting performance of the GARCH(1,1) model.

5. ACKNOWLEDGMENTS

The authors would like to thank Shiqing Ling and Peter Verhoeven for helpful discussions. The first author would like to acknowledge the Australian Research Council for an Ad Hoc PhD Scholarship and the Department of Economics at UWA for a C.A. Vargovic Memorial Fund Award. The second author wishes to acknowledge the financial support of the Australian Research Council.

6. REFERENCES

Bollerslev, T. Generalised autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327, 1986.

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

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

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

Franses, P.H. and D. van Dijk, Outlier detection in GARCH models, Econometric Institute Research Report 9926/A, Erasmus University Rotterdam, 1999.

He, C. and T. Teräsvirta, Properties of moments of a family of GARCH processes, *Journal of Econometrics*, 92, 173-192, 1999.

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

Ling, S. and M. McAleer, Stationarity and the existence of moments of a family of GARCH processes, to appear in *Journal of Econometrics*, 2001.

Verhoeven, P. and M. McAleer, Modelling outliers and extreme observations for ARMA-GARCH processes, Working Paper, Department of Economics, University of Western Australia, 2000.

Yew, X.C.H., M. McAleer and S. Ling, Determining an optimal window size for modelling volatility, to appear in A. Ullah, A. Wan and A. Chaturvedi (eds.), *Handbook of Applied Econometrics and Statistical Inference*, Marcel Dekker, New York, 2001.