

Modelling Time-Varying Volatility in Non-Ferrous Metals Markets

Clinton Watkins and Michael McAleer

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

Abstract Recently the modelling and forecasting of volatility has received much attention in the literature. As volatility is generally unobservable, it must be estimated. The GARCH(1,1) specification remains the most widely used time-varying financial volatility model in practice. This paper evaluates the adequacy and effectiveness of AR(1)-GARCH(1,1) in modelling and forecasting volatility in daily price returns on futures contracts for the two most important metals traded on the London Metal Exchange, namely aluminium and copper. The empirical analysis examines the properties of parameter estimates, robust t-ratios, moment conditions, and forecasts derived from rolling regressions.

Keywords: GARCH; Futures contracts; Volatility; Rolling regressions.

1. VOLATILITY AND METALS

Volatility in commodity markets represents risk to both 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.

This paper evaluates the adequacy and effectiveness of the AR(1)-GARCH(1,1) model in modelling and forecasting volatility in daily price returns on futures contracts for the two most important metals traded on the London Metal Exchange (LME), namely aluminium and copper. The analysis examines the properties of parameter estimates, robust t-ratios, moment conditions, and forecasts derived from a rolling regression model.

2. VOLATILITY MODEL

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

$$r_t = \mu + \phi r_{t-1} + \varepsilon_t, \quad |\phi| < 1, \quad (1)$$

and the conditional variance of ε_t is:

$$\varepsilon_t = \eta_t \sqrt{h_t}, \quad (2)$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \quad (3)$$

where r_t denotes returns on the futures price from period $t-1$ to t ; ε_t is the unconditional shock; the standardised 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, $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$ are sufficient conditions for the conditional variance to be positive. The ARCH coefficient, α , measures short run persistence in volatility, and the GARCH effect, β , measures the contribution to long run persistence, namely $\alpha + \beta$.

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 [2002] establish all the moment conditions for the general class of models, GARCH(p,q), and relate the moment conditions to the statistical properties of the models. The implications of the non-existence of moment conditions, such as the possible inconsistency of parameter estimates and invalid inference, are frequently ignored in the empirical literature on modelling volatility using GARCH processes.

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

$$\alpha + \beta < 1. \quad (4)$$

Under normality of ε_t , the fourth moment of the unconditional shock will exist if and only if the following condition is satisfied:

$$3\alpha^2 + 2\alpha\beta + \beta^2 < 1. \quad (5)$$

The second moment is sufficient for consistency of the (quasi-) maximum likelihood estimator, while the fourth moment is sufficient for asymptotic normality.

3. NON-FERROUS METALS DATA

Daily data for 3-month contract settlement prices are obtained from the LME for aluminium over the period 1 October 1982 to 31 August 2001, and for copper over the period 5 January 1976 to 31 August 2001. The aluminium price data set contains 4776

observations, and the copper data set contains 6474 observations. Prices quoted by the LME prior to July 1993 are denominated in Pound Sterling. The 3-month futures prices are converted from Pound Sterling to US Dollars using the 3-month US Dollar to Pound Sterling exchange rate. After July 1993, prices are quoted by the exchange in US Dollars. The returns series are calculated as:

$$r_{t-1,t} = (f_t - f_{t-1}) / f_{t-1} \quad (6)$$

Plots of the price and returns series for aluminium and copper are presented in Figures 1 and 2, respectively. Price series for both aluminium and copper show similar trends and structural breaks. Clusters of volatility are apparent in each of the two returns series. Returns are most volatile at clusters that coincide with the structural breaks in the price series. Distinct clusters of volatility occur at the structural break points.

Volatility in the returns of copper and aluminium are substantially different. While copper and aluminium sometimes share periods of clustered volatility at similar times, each market also contains periods of volatility not occurring in the other. Metals markets, while being affected by macroeconomic shocks, are also strongly influenced by market-specific events. The extent to which these permeate other markets depends on a number of factors, including the complimentary and substitute relationships between non-ferrous metals. However, it is unclear how these relationships between metals markets affect the extent to which shocks and periods of returns volatility occur in various markets. A particularly extreme example of a market-specific shock that is unique to a metals market was seen during 1996 in copper. The LME copper market was systematically manipulated by a trader in the Sumitomo Corporation of Japan during the early and mid 1990s (Gilbert 1996). Substantial volatility can be observed in returns around June 1996 (in the vicinity of observation 5151), when conditions in the copper market made Sumitomo's position untenable. At that time, hedge funds saw an opportunity to attack the inflated copper price, which fell from USD 2700 to USD 2000 per tonne over a four-week period.

Descriptive statistics for aluminium and copper returns are shown in Table 1. As is expected for a time series of returns, both the mean and median are very close to zero. Copper has a larger positive maximum daily percentage return, a lower negative minimum return, and a greater standard deviation than for aluminium returns. Copper and aluminium have essentially symmetric returns, with each displaying only a very small level of negative skewness. Both metal's returns are leptokurtic or fat-tailed, given their respective large kurtosis statistics in Table 1. The level of kurtosis in copper is larger than for aluminium. Both series are non-normal

according to the Jarque-Bera test, which rejects normality at the 1% level for each series.

Table 1: Descriptive statistics

Descriptive Statistic	Aluminium	Copper
Mean	7.92E-05	3.40E-05
Median	0.0000	0.0002
Maximum	0.0733	0.0972
Minimum	-0.1384	-0.1643
Std. Dev.	0.0131	0.0147
Skewness	-0.2637	-0.4070
Kurtosis	9.2660	9.7535
Jarque-Bera ¹	7866.89 (0.00)	12479.94 (0.00)

1. Prob-values in parentheses.

4. MODELLING AND FORECASTING

The AR(1)-GARCH(1,1) model is estimated for each returns series using a rolling window of 750 observations, which rolls 4026 times over a sample of 4775 returns observations for aluminium, and 5724 times over a sample of 6437 returns observations for copper. Estimates from the rolling samples are treated as "data" in the descriptive discussion below. Each model is estimated by maximum likelihood. The forecast volatility is compared with the 'true' volatility calculated over the same window, the 'true' volatility being defined as:

$$v_t = (r_t - \bar{r})^2 \quad (7)$$

where v_t refers to the 'true' volatility at time t , and \bar{r} is the mean return over the window for the sample used. The 1-day ahead forecast error, u_{t+1} , is defined as:

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

4.1 Parameter Estimates

Plots of the rolling α coefficient estimates are provided in Figure 3 for aluminium and Figure 4 for copper. Rolling β estimates are shown in Figure 5 for aluminium and Figure 6 for copper. In almost all cases, the parameter estimates are positive, as required for the GARCH(1,1) model. Exceptions are three estimates for the copper α coefficient, namely small and negative occurring in rolling windows 2966, 2967 and 3046. As expected, the majority of α parameter estimates are small at around 0.1, while most β parameter estimates are large at around 0.9.

However, there are numerous instances where the parameter estimates depart substantially from their typical values, either in the form of a level shift, period of variability, or a one-off extreme spike in the estimate. Furthermore, the rolling estimates for each metal appear fundamentally different.

Rolling α estimates for aluminium are initially at a level typically expected for short run persistence. The

estimates are around 0.1 over the first 1595 rolling windows, and trend slightly downwards. After a turning point in window 1594 the estimates begin a steep upward trend, reaching a level of 0.3, which is unusual for an α estimate. A second turning point occurs at window 1878, and subsequently the estimates fall to a level below 0.1. During this downward trend, there are several upward and downward spikes in the estimated α . Within 50 rolling windows, the estimate varies between 0.35 and 0.01. Between rolling windows 2370 and 3109, the estimate is around 0.1, but beyond 3110 to 3536, it trends downward to around 0.01. After window 3537, the trend becomes positive and the estimate climbs to above 0.1. During this upward trend, the estimate is highly variable between windows 3737 and 3882, with several single estimated downward spikes to around 0.01 in α . While the trend is towards an estimate of 0.1, numerous estimates are close to zero, indicating no short run persistence in volatility.

The pattern in the plot of rolling estimates of α for copper is markedly different to that for aluminium, bearing in mind that the sample is also longer. Estimates of α for copper returns trend upward from 0.1 from the start of the sample until a level shift in the estimates occurs at window 1944. The estimates of α fall to a level of around 0.3 until window 3288. During the period between windows 2910 and 3288, the estimates are variable, with three being negative. In such cases, the GARCH process may not exist, as the conditional variance cannot be negative. Beyond window 3289 an upward trend in the estimates emerges and continues until window 3907. After window 3907, the α estimates trend downward from around 0.1 to 0.03 at the end of the sample. This trend involves two small level shifts, one upward at window 4390 and one downward at window 5128, and a large spike in the estimate during estimation window 4815. An α estimate of 0.297 is generated from window 4815, which is far greater than neighbouring estimates, which are 0.077 and 0.059.

The α estimates for the aluminium and copper returns series indicate that short run persistence varies in importance over each respective sample. Substantial variability in the estimate for each series is typical. A number of the features of the rolling α estimates coincide with shocks to the returns series.

Estimates for the β parameter appear in Figures 5 and 6 for aluminium and copper returns, respectively. The majority of rolling estimates for both metals series are greater than 0.8 and less than 1, as is to be expected. However, as in the case for the α estimates, the rolling estimates for aluminium and copper have quite different characteristics over time.

The rolling β estimates for aluminium returns show a steep downward trend from around 0.9 to almost 0.3

over the period from window 1534 to 2048. From window 2049 to 2273, the rolling estimates return to their former values. However, during this period several extremely large positive and negative spikes occur in the value of the estimate, especially in the neighbourhood of window 2196. Eight estimates are below 0.25, while several high positive spikes also occur, the highest being above 0.8 in window 2187. Estimates of β remain in the typical region of 0.8 to 0.9 until window 3722, beyond which a period of extreme variability occurs, and negative estimates are observed. Estimates of β between -0.185 and 0.952 are generated from the rolling regressions. Beyond window 3849, numerous estimates are negative. The last 212 estimates settle close to zero, indicating there is no contribution to long run persistence in volatility that can be detected by the model.

The majority of the rolling estimates for β from the copper data are between 0.8 and 1, with one estimate being greater than 1 in window 3046. Relatively few estimates are below 0.6 when compared with the corresponding aluminium β estimates. No estimates for copper returns are negative. A downward trend in the β estimates starts at window 1019 and continues until 1952, when a level shift moves the estimate to 0.975, and beyond that the estimate remains between 0.8 and 1 for most of the sample. A notable feature of the copper β estimates beyond window 1952 is the presence of negative spikes which last for one window. The three largest of these occur in windows 4815, 5172 and 5176. In addition, there are numerous small negative spikes over the entire estimation period. Small positive spikes occur frequently between windows 2933 and 3257.

4.2 Robust Rolling t-ratios

For each set of rolling α and β estimates, rolling robust t-ratios are generated, plots of which are shown in Figures 7, 8, 9 and 10. Each plot shows that the rolling robust t-ratios vary considerably over the sample. For all except the copper β estimate, there are periods in which the estimates are not significant.

Figures 7 and 8 show the robust t-ratios for the α estimates derived from the aluminium data and copper data, respectively. The t-ratios for the aluminium data show an obvious and substantial downward trend over the sample, while those for copper show only a slight trend. Almost all the aluminium α estimates are significant for the first 3179 rolling windows. The rolling estimates near the end of the sample have variable and low t-ratios, which indicate most of the α estimates over approximately the last 850 windows are not significant.

Copper t-ratios for α remain in the vicinity of 2 for much of the sample, but become variable in the middle of the sample between windows 1850 and

3542. Within this period, t-ratios can be as high as 6.256 and as low as -1.942, but the majority of the estimates are significant. Interestingly, this period of variability in the t-ratios coincides with the level shift in the copper α estimates which occurs at window 1944, and a period that contains numerous small positive and negative spikes in the α estimates.

Figures 9 and 10 show the t-ratios for the aluminium and copper β estimates, respectively. Typically, the t-ratios for both series are large, when the estimates are above 0.8 and less than 1. Where the aluminium β estimate trends downward from window 1534 to 2048, the t-ratio also trends downward. Between windows 2071 and 2595, both the β estimate and its t-ratio trend upward. Toward the end of the sample, after window 3722 when the aluminium β becomes extremely variable, the t-ratio becomes low and most estimates are not significant.

On the whole, the t-ratios for the copper β estimates are large. As with the α estimates, the period after the level shift that occurs in both the α and β estimates, between windows 1953 and 3542, sees the t-ratio for the β estimate increase substantially. The t-ratios also become variable, and contain frequent large negative and positive spikes, but the estimate always remains significant. The three large negative spikes in the β estimate for copper returns in windows 4815, 5172 and 5176 are accompanied by downward spikes in the robust t-ratio to values which indicate the estimates are not significant. Large upward spikes in the robust t-ratios occur for four β estimates that are very close to 1, namely those for windows 5137, 5146, 5147, and 5153. The estimate contains several small negative spikes after window 5176, and the t-ratio varies over this period until the end of the sample.

In general, unusually high estimates for α and unusually low estimates for β are frequently associated with low robust t-ratios. Spikes in parameter estimates often cause variable or volatile t-ratios, but these spikes do not always lead to low t-ratios. Large changes or level shifts in estimates are associated with volatile t-ratios.

4.3 Rolling Moment Conditions

Both the second and fourth moment conditions are satisfied more frequently in models for copper than for the aluminium returns data. Almost 15% of the models do not satisfy the fourth moment condition and 13% of the models do not satisfy the second moment for aluminium returns. For copper, the failure rate for the regularity conditions is approximately 12% and 9% for the fourth and second moments, respectively.

Plots of the second and fourth moments of model for aluminium returns volatility are provided in Figures

11 and 12, respectively. Each moment exceeds 1, between windows 526 and 1038. This corresponds to rolling windows for about two years of trading days, so that the moment conditions are violated for a substantial period in the sample. Furthermore, the fourth moment condition is also violated between windows 492 and 526, 1038 and 1051, and 1178 to 1250. In addition, the second moment exceeds 1 between windows 1217 and 1224.

The rolling second and fourth moments in Figures 13 and 14, respectively, show quite a different pattern for copper. For the majority of rolling windows, both moments are between 0.9 and 1. Deviations from this range are less frequent, and lower in magnitude than for aluminium. Neither the second nor fourth moment is negative at any point during the sample. However, there are several periods in which estimation windows result in second and fourth moments, or only fourth moments, for the GARCH(1,1) model equal to or greater than 1. Most notable is a period of over two years of trading days from window 2224 to 2709, in which both the second and fourth moments exceed 1. The second moment condition is violated for windows 2027, 2029, and for 22 windows between windows 3219 and 3386. The fourth moment condition is violated frequently between windows 274 and 439, for windows 2027 and 2029, and for 88 windows between windows 3218 and 3655. In total, the moment conditions are satisfied more frequently for the rolling GARCH model of copper returns volatility than for aluminium.

Volatility persistence for copper is very close to 1 for the entire sample, except for the negative spikes and one level shift which remains for 282 windows, or just over one year between windows 1670 and 1952, before returning to previous levels.

4.4 Rolling One-step-ahead Forecasts

Forecasts of volatility generated by the models are compared using 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). Table 2 shows the forecast performance of the models for aluminium and copper returns.

The forecast performance of the models for aluminium and copper is similar in terms of ME, MAE, MedAE, RMSE and RMedSE. For forecasts of volatility in both metals markets, median errors are always smaller than the comparable mean errors. In general, forecast errors are slightly greater for copper volatility forecasts than for aluminium. When RMSE calculated separately for positive and negative forecast errors, RMSE(-) is substantially larger than

RMSE(+)) for both markets. Volatility appears to be under-forecast by GARCH(1,1).

Table 2: Forecast evaluation criteria

Evaluation Criteria	Aluminium	Copper
ME	0.00343	0.00397
MAE	0.00716	0.00824
MedAE	0.00625	0.00708
RMSE	0.00946	0.01091
RMSE(-)	0.01169	0.01305
RMSE(+)	0.00859	0.01005
RMedSE	0.00625	0.00708
SMAPE	75.91	75.09
SMedAPE	64.84	63.49
SMWAPE	51.84	51.22
SMedWAPE	32.49	33.08
R ²	0.15	0.12
R ² _e	0.66	0.69
% Forecasts Under	24.89	26.19
% Forecasts Over	75.11	73.81

The forecast performance of the GARCH(1,1) model is similar for aluminium and copper returns volatility according to SMAPE, SMedAPE, SMWAPE and SMedWAPE. Aluminium percentage forecast errors are slightly higher than for copper when evaluated using SMAPE, SMedAPE and SMWAPE, but the opposite is true under SMedWAPE. Again, the median measures are substantially lower than their mean counterparts. The weighted forecast performance measures, SMWAPE and SMedWAPE, are considerably lower, by around one-third, than their non-weighted counterparts. This suggests that a large number of forecast errors occurs when the actual volatility of the forecast period is lower than the average volatility in the entire sample.

In Table 2, R^2 is obtained by regressing ex-post volatility on forecast volatility. For both the aluminium and copper forecasts, the R^2 values are relatively low at 0.15 and 0.12, respectively, indicating that the predictive power of the GARCH(1,1) model is poor. R^2_e is the coefficient of determination by regressing the forecast errors on the ex-post volatility. Values for this measure are greater at 0.66 for aluminium volatility and 0.69 for copper. In both markets, the actual volatility has a high degree of explanatory power over forecast errors, thereby indicating that the GARCH(1,1) model does not generate good forecasts. Anderson and Bollerslev (1998) argue that because noise in daily actual volatilities results in poor predictive power for the GARCH(1,1) model, that higher frequency data can improve the intertemporal stability of volatility, and also improve the explanatory and forecasting power of the model.

In the case of aluminium returns volatility, the forecasting model over-predicts volatility around 75% of the time, and under-predicts around 25% of the one-step ahead forecasts. A similar situation

exists for the model forecasting copper returns volatility. Almost 74% of the time, the model over-forecasts, while around 26% of the forecasts under-predict actual volatility.

Rolling one-step-ahead volatility forecasts for aluminium futures return volatility from AR(1)-GARCH(1,1) are shown in Figure 15. Forecasts from the model capture the major features of the actual volatility in aluminium futures returns over the sample. The most notable cluster of volatility in the forecasts is evident between 21 October 1987 and 25 May 1990, referring to the forecasts for observations 1276 to 1678. This period coincides with the largest negative return entering the estimation window, that for 20 October 1987 (observation 1275), followed by numerous large negative and positive returns.

Forecasts of high volatility also occur in December 1985 (forecasts for observations 814 to 816), early September 1987 (forecasts for observations 1242 and 1243), October 1990 (forecasts for observations 2020 to 2022), October 1991 to January 1992 (forecasts for observations 2284 to 2347), and February to March 1995 (forecasts for observations 3120 to 3139). While the model fails to forecast shocks such as the large negative return in the market on 20 October 1987, as is to be expected, forecasts immediately after the large shock over-estimate volatility persistence. This is particularly evident in the estimates produced by windows 527 to 770, where large positive forecast errors are apparent.

A plot of the rolling volatility forecasts for copper futures returns is given in Figure 16. The forecasts generally appear similar to the profile of actual volatility in returns. Several volatility clusters are clearly indicated. The most obvious of these corresponds to the October 1987 stock market crash. When the large negative return corresponding to 20 October 1987 (observation 2973) enters the estimation windows (beginning in window 2224), there is a spike in the volatility forecasts produced. High levels of volatility persist in the forecasts, as a result of the 20 October 1987 return, as well as further large negative and positive returns, until window 3143, generating a cluster of high forecasts that dominates Figure 16.

Surprisingly, the events of October 1987 do not give rise to the highest forecasts. The greatest forecast volatility occurs during the Sumitomo collapse. The first large negative return during this period, corresponding to 20 May 1996, enters the sample in window 4390. A substantial spike in forecast volatility is generated. The highest forecast of 0.0475 for observation 5153 is generated when the large negative return of 6 June and large positive correction of 7 June 1996 (observations 5151 and 5152 respectively) enter the estimation window. The cluster of volatility forecasts persists from windows

4390 to 4424, corresponding to forecasts for 21 May to 6 July 1996 (forecasts for observations 5140 to 5174). Frequent and large negative forecast errors observed for both metals can be interpreted as an inability of the GARCH(1,1) model to forecast volatility well in the tails of the returns distribution.

5. CONCLUSION

This paper has evaluated the performance of the AR(1)-GARCH(1,1) model in modelling and forecasting daily returns volatility in two futures markets for industrially-used non-ferrous metals traded on the London Metal Exchange, namely aluminium and copper. Rolling parameter estimates, robust t-ratios, moment conditions, one-step-ahead forecasts and forecast evaluation criteria were analysed.

ACKNOWLEDGEMENTS

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.

REFERENCES

- Anderson, T.G. and T. Bollerslev, Answering the sceptics: yes, standard volatility models do provide accurate forecasts, *International Economic Review*, 39, 885-905, 1998.
- Bollerslev, T. Generalised autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307-327, 1986.
- Gilbert, C.L., Manipulation of metals futures: lessons learned from Sumitomo, Discussion Paper No. 1537, Centre for Economic Policy Research, 1996.
- Ling, S. and M. McAleer, Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models, *Econometric Theory*, 18, 722-729, 2002.

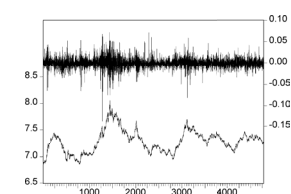


Figure 1: Logarithm of aluminium prices (bottom) and returns (top)

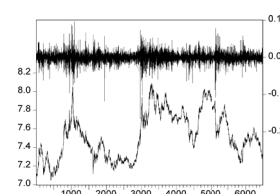


Figure 2: Logarithm of copper prices (bottom) and returns (top)

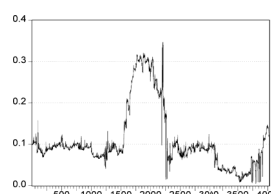


Figure 3: Aluminium α estimates

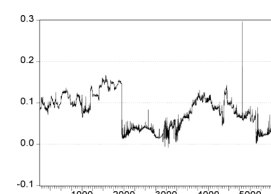


Figure 4: Copper α estimates

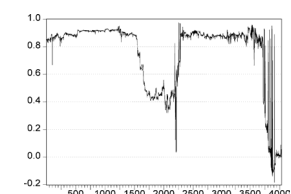


Figure 5: Aluminium β estimates



Figure 6: Copper β estimates



Figure 7: Aluminium α estimate robust t-ratios

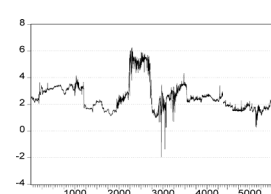


Figure 8: Copper α estimate robust t-ratios

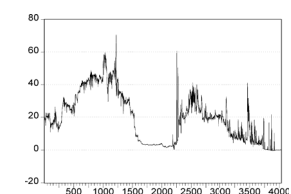


Figure 9: Aluminium β estimate robust t-ratios

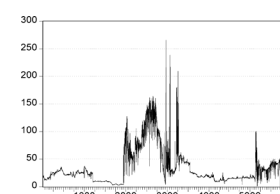


Figure 10: Copper β estimate robust t-ratios

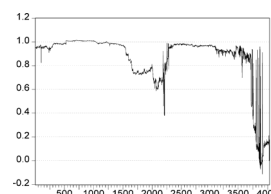


Figure 11: Aluminium second moment

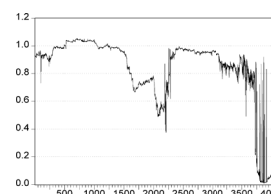


Figure 12: Aluminium fourth moment

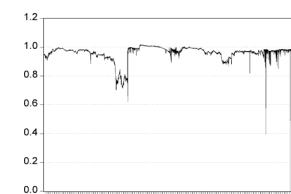


Figure 13: Copper second moment

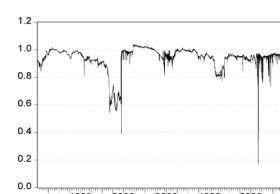


Figure 14: Copper fourth moment

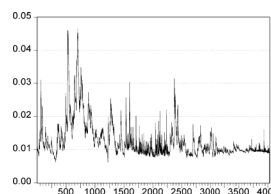


Figure 15: Aluminium forecast volatility

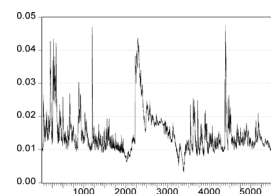


Figure 16: Copper forecast volatility