Outliers in Time Series

Vlad Ardelean

Department of Statistics and Econometrics, University of Erlangen-Nuremberg

26.08.2011

Outline

Motivation

Analysis of financial time series

In parametric time series analysis there is the implicit assumption that there are no outliers.

An outlier is an observation that deviates much from other observations. It is likely that it was 'generated' by a different process.

There are different reasons for the identification of outliers.

- The outlier is to be rejected.
- There is a special interest (detecting alternative or rare phenomena) in that observation.
- Outliers can be used as diagnostic indicators.

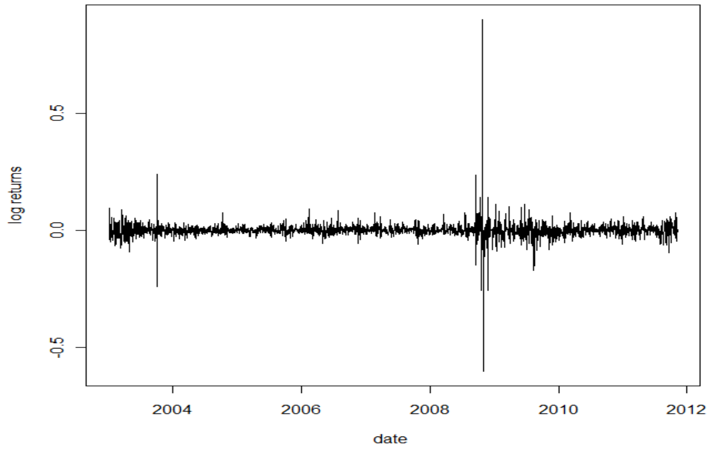

Figure: Log returns of VW Stock

GARCH processes

GARCH processes

A stochastic process X_t is a **GARCH(p,q)** process, cf. Bollerslev (1986), if:

$$X_t|\mathcal{F}_{t-1} = \sigma_t \nu_t,$$

$$\sigma_t^2 = (\sigma_t(\gamma))^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \; t \in \mathbb{Z}$$

with $\gamma = (\alpha_0, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q)$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, p$ and $\beta_i \geq 0$, $i = 1, \ldots, q$.
Furthermore $\mathcal{F}_t$ denotes the information set of the process up to time t.
The innovations $\nu_t \overset{iid}{\sim} G$, where G is some distribution function with $E_G(\nu_t) = 0$ and $E_G(\nu_t^2) = 1$.
The log likelihood (for normal innovations) is given by (apart from constants):

$$LogL(\gamma) = \ell(\gamma) = \sum_{i=1}^{n} \frac{1}{2} \log(\sigma_t) + \frac{x_t^2}{\sigma_t^2}$$

Detecting Outlying Observations

There are different strategies to detect outlying observations, including:

- robust estimators,
- likelihood ratio test and
- tests based on the cumulative sum of observed residuals,

Outlier in GARCH Processes

Two types of outlier exist, additive and innovational outliers, cf. Fox (1972). Additive outliers only influence one period, while innovational outliers influence more than one period. Following Doornik and Ooms (2005) they can be modelled the following way:

- Additive outlier:

$$
\begin{aligned}
Y_t &= X_t + \gamma 1_t(\tau) \\
X_t|\mathcal{F}_{t-1} &\sim N(0, \sigma_{t-1}^2), \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2,
\end{aligned}
$$

- Innovational outlier:

$$
\begin{aligned}
Y_t &= X_t + \gamma 1_t(\tau) \\
X_t|\mathcal{F}_{t-1} &\sim N(0, \sigma_{t-1}^2), \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2,
\end{aligned}
$$

Likeliood ratio test

Let $y_1, \cdots, y_n$ be the realisations of the observed process and let $x_1, \cdots, x_n$ be the realisations of the unobservable process

$$
\begin{aligned}
y_t &= \kappa_1 1_t(\tau) + x_t \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 + \sum_{i=1}^{p} \kappa_{1+i} 1_t(\tau - i),
\end{aligned}
$$

$$\lambda_\tau = -2(\log L(\hat\gamma_0) - \log L(\hat\gamma_1)) \overset{a}{\sim} \chi^2(p+1),$$

where, $\hat\gamma_0 = (\hat\alpha_0, \hat\alpha_1, \ldots, \hat\alpha_p, \hat\beta_1, \ldots, \hat\beta_q)$ is the restricted ML-estimate and $\hat\gamma_1 = (\hat\alpha_0, \hat\alpha_1, \ldots, \hat\alpha_p, \hat\beta_1, \ldots, \hat\beta_q, \hat\kappa_1, \ldots, \hat\kappa_{p+1})$ is the unrestricted ML-estimate. Since the time of an outlier is unknown, the test statistic is computed for every $\tau = 1, \ldots, T$

$$M_n = \max_{t \le n} \lambda_t.$$

Testing for structural breaks

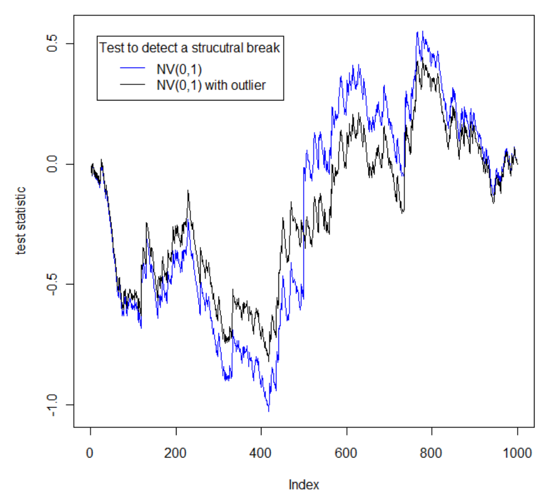

Figure: Testing for structural breaks Inclan and Tiao (1994)

A Test based on the cumulative Sum

Theorem (Theorem 11 from Merlevède et al. (2006))

Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary sequence with $E(X_0)=0$ and $E(X_0)^2<\infty$ Assume that the following holds:

$$\sum_{i=1}^{\infty}\frac{||E(S_n|\mathcal{F}_0)||_2}{n^{\frac{3}{2}}}<\infty,$$

where $S_n=\sum_{i=1}^{n}X_i$ and $||X||_p=(E(|X|^p))^{\frac{1}{p}}$. Then,

$$\left\{max_{1\leq k\leq n}\frac{S_k^2}{n}\ : n\geq 1\right\}$$

is uniformly integrable and

$$W_n\xrightarrow{D}\sqrt{\eta}W,$$

where $W_n(r)=\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{\lfloor nr\rfloor}X_i$ and W a Brownian Motion, η is a non-negative random variable with finite mean $E[\eta]=\sigma^2$ and independent of $\{W(r);r\geq 0\}$.

Theorem

Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary process that fulfils the assumptions of the previous theorem and let $\xi_t=X_t^2-Var(X_t)$, then

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{\lfloor rT\rfloor}\xi_i\xrightarrow{D}W(r),$$

where $\sigma=E\left((\sum_{i=1}^{n}X_i^2/n)^2\right)$.
It holds furthermore that:

$$\max_{1\leq i\leq n}\xi_i-\xi_{i-1}\rightarrow G,$$

where G is a Gumbel distribution with suitable normalizing constants $\mu_g(n)$ and $\sigma_g(n)$.

Proof (Theorem 2) From Theorem 1 we have that

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{\lfloor rn\rfloor}\xi_i\xrightarrow{D}W(r).$$

By definition the following holds for a Brownian motion:

- W(0)=0.
- Let $t_1,t_2,t_3,t_4\in[0,1]$ with $t_1<t_2,t_3<t_4$. Then $W(t_2)-W(t_1)$ and $W(t_4)-W(t_3)$ are stochastically independent.
- $\forall\ t_1,t_2\in[0,1]$ with $t_1\leq t_2$ it holds: $W(t_2)-W(t_1)\ N(0,t_2-t_1)$.

From this follows that $\xi_t-\xi_{t-1}$ is i.i.d. normal with $\mu=0$ and $\sigma^2=\frac{1}{n}$
Furthermore, the maximum of i.i.d. normal distributed random variables lies in the domain of attraction of a Gumbel distribution. The location parameter $\mu_g(n)$ and the scale parameter $\sigma_g(n)$ are given in Takahashi (1987)

$$\mu_g(n)=\left((2\log(n))^{\frac{1}{2}}-\left(\log(\log(n)+\log(4\pi)\right)/\left(2/(2\log(n))^{\frac{1}{2}}\right)\right)\sqrt{\frac{1}{n}}\qquad(1)$$

$$\sigma_g(n)=\left(2\log(n)\right)^{-\frac{1}{2}}\sqrt{\frac{1}{n}}\qquad(2)$$

Simulation study

- The test based on the increments of the Brownian motion and the test based on the likelihood ratio test are compared.
- 500 repetitions
- GARCH(1,1) with $\alpha_0 = 0.001, \alpha_1 = 0.1, \beta_1 = 0.8$, $n = 500$ and $= 1000$
- Relative and fixed outliers at time $\tau = n/2$:
- additive and innovational outliers of size $3\sigma_t$ and $5\sigma_t$.
- $$\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - \hat{\sigma}^2\right)^2 + \frac{2}{n}\sum_{j=1}^{m} w(l,m) \sum_{i=j+1}^{n}\left(x_i^2 - \hat{\sigma}^2\right)\left(x_{i-j}^2 - \hat{\sigma}^2\right),$$

  where $w(l,m)$ is a lag window, i.e. the Bartlett window defined by

  $$w(j,m) = 1 - \frac{j}{m+1}.$$

Observations	LR		$CUSUM - type$	
	0.95	0.99	0.95	0.99
n=500	0.086	0.016	0.066	0.014
n=1000	0.068	0.012	0.058	0.012

	Obs	size	LR		CUSUM	
			0.05	0.01	0.05	0.01
Rel add	500	3	0.4 (0.332)	0.212 (0.188)	0.298 (0.27)	0.222 (0.21)
		5	0.968 (0.95)	0.944 (0.93)	0.92 (0.92)	0.864 (0.864)
	1000	3	0.442 (0.342)	0.184 (0.172)	0.346 (0.258)	0.258 (0.2)
		5	0.978 (0.966)	0.954 (0.946)	0.96 (0.952)	0.928 (0.922)
Fixed add	500	3	0.454 (0.402)	0.24 (0.22)	0.322 (0.278)	0.238 (0.214)
		5	0.956 (0.948)	0.908 (0.902)	0.944 (0.944)	0.878 (0.878)
	1000	3	0.472 (0.388)	0.226 (0.21)	0.318 (0.252)	0.234 (0.19)
		5	0.974 (0.962)	0.928 (0.92)	0.97 (0.964)	0.962 (0.956)
Rel innov	500	3	0.34 (0.258)	0.144 (0.128)	0.198 (0.17)	0.124 (0.116)
		5	0.996 (0.966)	0.976 (0.948)	0.756 (0.754)	0.612 (0.612)
	1000	3	0.388 (0.308)	0.208 (0.186)	0.252 (0.182)	0.216 (0.162)
		5	0.996 (0.978)	0.984 (0.97)	0.874 (0.858)	0.808 (0.794)
Fixed innov	500	3	0.372 (0.322)	0.202 (0.184)	0.236 (0.198)	0.156 (0.134)
		5	0.964 (0.942)	0.904 (0.894)	0.754 (0.752)	0.634 (0.634)
	1000	3	0.428 (0.344)	0.212 (0.19)	0.266 (0.208)	0.196 (0.158)
		5	0.98 (0.97)	0.94 (0.934)	0.91 (0.898)	0.832 (0.824)

Table: Power of the LR-test and the CUSUM-test

Summary and Outlook

- The proposed method to detect outliers has a similar power as the likelihood ratio test. A great advantage is the low computational cost.
- The high runtime for the likelihood ratio test is due to the fact that for every observation the maximum likelihood has to be computed. In order to reduce the observations that are possible candidates for outliers, a model free variance can be computed Gelper et al. (2009)
- The proposed method to detect outliers can be extended to multivariate processes, especially:
  - Vector autoregressive moving-average (VARMA) processes
  - Multivariate GARCH processes (CCC-GARCH)

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–322.

Doornik, J. A. and Ooms, M. (2005). Outlier Detection in GARCH Models.

Fox, A. J. (1972). Outliers in Time Series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(3):350–363.

Gelper, S., Schettlinger, K., Croux, C., and Gather, U. (2009). Robust online scale estimation in time series: A model-free approach. *Journal of Statistical Planning and Inference*, 139(2):335–349.

Inclan, C. and Tiao, G. C. (1994). Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance. *Journal of the American Statistical Association*, 89(427):913–923.

Merlevède, F., Peligrad, M., and Utev, S. (2006). Recent advances in invariance principles for stationary sequences. *Probability Surveys*, 3:1–35.

Takahashi, R. (1987). Normalizing constants of a distribution which belongs to the domain of attraction of the Gumbel distribution. *Statistics & Probability Letters*, 5(3):197–200.