Exercise GARCH
Principles of Programming in Econometrics

MSc/PhD Bootcamp
Vrije Universiteit Amsterdam & Tinbergen Institute
C.S. Bos, `c.s.bos@vu.nl`

The exercise below is an exercise, to be solved over the course of the next weeks.

1 Setting and model

Your exercise for this exam is to estimate a so-called GARCH model (Engle, 1982; Bollerslev, 1986) on inflation data y_t. The basic model reads

$$y_t = \mu_t + a_t, \qquad a_t \sim \mathcal{N}(0, \sigma_t^2), \tag{1}$$

$$\mu_t = X_t \beta, \tag{2}$$

$$\sigma_{t+1}^2 = \omega + \alpha a_t^2 + \delta \sigma_t^2, \tag{3}$$

$$\sigma_1^2 \equiv \frac{\omega}{1 - \alpha - \delta}, \qquad t = 1, \ldots, T. \tag{4}$$

Equation (1) is a standard regression equation, with a time varying variance. The loglikelihood for observation y_t is like the loglikelihood of the normal model with mean μ_t and variance σ_t^2. The mean μ_t is specified in (2) like a standard regression, with k explanatory variables in X_t, linking with parameters β.

The added value of the GARCH model is the variance equation (3). The variance slowly evolves over time, where a large shock a_t^2 leads to a sudden increase in the next-period variance σ_{t+1}^2. Initialisation of the first variance is done according to (4).

The variance σ_t^2 of course needs to stay positive. A sufficient condition for this is to ensure that the parameters ω, α, δ are non-negative. Furthermore, e.g. if $\delta > 1$ or $\alpha > 1$ you can see that the variance sequence becomes explosive. Hence, a further restriction is that $\alpha + \delta < 1$.

Note that, given a vector of parameters $\theta = (\beta', \omega, \alpha, \delta)'$ and the data $Y = (y_1, \ldots, y_T)$, a full vector of $\sigma^2(\theta, Y) = (\sigma_1^2, \ldots, \sigma_T^2)$ can be extracted.

2 Data

Originally the ARCH effect was found in inflation data (Engle, 1982). To check on these results, use the inflation data constructed from the price index `CUUR0000SA0` (source: Bureau of Labor Statistics), over the period 1958–2018 (or more recent, download yourself). Provided is a data file with the price index, say P_t; construct the percentage inflation rate through

$$y_t = 100 \left(\log(P_t) - \log(P_{t-1}) \right).$$

As explanatory variables, use a $T \times 17$ matrix X consisting of

- a constant,
- 11 seasonal dummies $S_{ti} = I_{m_t = i}, i = 1, \ldots, 11$,
- 5 level shift dummies $D_{ti} = I_{t \geq d_i}$, with $d_i \in \{1973{:}7, 1976{:}7, 1979{:}1, 1982{:}7, 1990{:}1\}$.

3 On the restrictions

When estimating the GARCH model, one has to ensure that the restrictions on the parameters are satisfied. You can do this either using the SQP approach, or using the transformation approach. With the latter option, it is relatively hard to implement the restriction that $\alpha + \delta < 1$. In general it appears to be sufficient to impose $0 < \omega, 0 < \alpha < 1, 0 < \delta < 1$, and only indicate failure (returning a 0) for computing the loglikelihood if the sum of α and δ happens to be too large.

4 Questions

Basically, estimate the model, give parameter output including standard deviations, and the optimal loglikelihood that you found. Preferably add the number of function evaluations that you needed for obtaining the optimum. Over the period 1958-present, do this for

- the pure regression model (without GARCH effects)
- the pure GARCH model (without mean effect, set $\mu_t = \beta_0$ constant)
- and the full GARCH-M model, estimating the full $\theta = (\beta', \omega, \alpha, \delta)'$.

What model do you prefer?

Also, give some graphical output. E.g. plot the inflation y_t together with the estimated mean process μ_t, and plot the standard deviation σ_t or variance σ_t^2. Compare density plots of $y_t, a_t = y_t - \mu_t$ and $\epsilon_t = a_t/\sigma_t$. Which one seems normally distributed?

If you hand in, then please prepare a very short report (max. 5 pages excluding graphs/tables, probably shorter is already sufficient) where you discuss your findings, possibly the problems you encountered, and indicate if there is something you might want comments on.

Zip together a PDF version of the report, plus ready-to-run Python programs and the data; if necessary add a `readme.txt` if I would have to run the programs in a specific order. Hand it in through the Canvas page of the course.

5 Help

This project is seems relatively large, especially if you don't start it in a structured fashion. Check clearly, before you start, how to split it up in more manageable subtasks. Large parts of the project you should have done during the course already.

It can be very helpful to first generate data from a pure GARCH model, using e.g. $\theta = (\beta', \omega, \alpha, \delta)' = (1, .05, .05, .9)'$, and estimate the generated data. Note that the data should have mean $\mu = 1$ and unconditional variance $\sigma^2 = \mathrm{E}\,\sigma_t^2 = \omega/(1 - \alpha - \delta) = 1$. If you generate $T = 5000$ observations and re-estimate the parameters, you should be able to get rather close to the parameters of the DGP.

References

Bollerslev, Tim (1986). “Generalized Autoregressive Conditional Heteroskedasticity”. In: *Journal of Econometrics* 31.3, pp. 307–327. DOI: `10.1016/0304-4076(86)90063-1`.

Engle, Robert F. (1982). “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”. In: *Econometrica* 50, pp. 987–1008. DOI: `10.2307/1912773`.