Adrian Beer

1197304

## Task B.1

Let $\{x_t\}_{t\in\{1,...,T\}}$ be the returns of the Apple stocks.

Assuming they follow a AR(1)−T−GARCH(2,2) the data generating process is

$$x_{t+1} = \hat{x}_{t+1} + \varepsilon_{t+1}$$

where we model $\hat{x}_{t+1} := \phi_0 + \phi_1 x_t$, as the conditional mean of $x_{t+1}$ $\mathbb{E}[x_{t+1} | \mathcal{F}_t]$ as an AR(1)-process. $\mathcal{F}_t$ is the information set available at time t.

We model the conditional variance of $\varepsilon_{t+1}$, $\mathbb{V}(\varepsilon_{t+1} | \mathcal{F}_t)$ with the T−GARCH(2,2), i.e.

$$\varepsilon_{t+1} = \hat{\sigma}_{t+1} \eta_{t+1}$$ where $\eta_{t+1}$ follows a white noise process, i.e. WNP(0,1) with variance of 1 and mean 0 and ~~...~~.

$$\hat{\sigma}^2_{t+1} = \mathbb{V}(\varepsilon_{t+1} | \mathcal{F}_t) =$$
$$\alpha_0 + (\alpha_1 + \theta_1 \mathbb{1}_{(\varepsilon_t < 0)}) \cdot \varepsilon_t^2 + (\alpha_2 + \theta_2 \mathbb{1}_{(\varepsilon_{t-1} < 0)}) \varepsilon_{t-1}^2 \quad (*)$$
$$+ \beta_1 \hat{\sigma}_t^2 + \beta_2 \hat{\sigma}_{t-1}^2.$$

$\mathbb{1}$ is the indicator function: $\mathbb{1}_{(\varepsilon_t < 0)} \begin{cases} 1 & \varepsilon_t < 0, \\ 0 & \text{sonst.} \end{cases}$

We can assume that the ~~mean~~ cond. mean of $\varepsilon_{t+1}$ is 0, because of the AR(1) term $\hat{x}_{t+1}$.

Adrian Beer
1997304

Now for the joint log-likelihood function:

We need to assume some distribution of $\eta$. Let $f$ be the probability density function of $\eta$. It must be a WNP(0,1).

The following describes the log-likelihood function $\ell(\lambda|X)$ where $X = \{x_t\}_{t \in \{1,..,T\}}$ and $\lambda$ is the parameter vector which we want to optimize with a maximum likelihood estimation (MLE). The parameter vector $\lambda$ consists of $(\phi_0, \phi_1, \theta_1, \theta_2, \alpha_1, \alpha_2, \beta_1, \beta_2, \hat{\sigma}_2, \hat{\sigma}_3)$ — these are needed because of (*)

$$\hat{\{x_t\}}_t \quad \{\hat{x}_t\}_{t \in \{2,..,T\}} = \phi_0 + \phi_1 \{x_t\}_{t \in \{1,..,T-1\}} ,$$

so in vector notation $\hat{X} = \phi_0 + \phi_1 X_{[1:T-1]}$ ,

$(T-1 \times 1)$ $(T-1 \times 1)$ scalar $(T-1 \times 1)$

We can now obtain $\varepsilon = \hat{X} - X_{[1:T-1]} = \{\varepsilon_t\}_{t \in \{2,..,T\}}$

For the T-GARCH we need a for loop as well as initial estimates for $\hat{\sigma}_2, \hat{\sigma}_3$. (*)

for $t = 4,...,T$ (the first two epsilons are available at $t=2, t=3$ only and they are required for the calculation)

$$\hat{\sigma}^2_{t+1} = \quad ... \quad \text{see} \quad (*)$$

Then $\ell(\lambda|X) = \sum_{t=4}^{T} \log f\left(\frac{x_t - \hat{x}_t}{\hat{\sigma}_t}\right)$ is the log likelihood function.

We can then maximize $\ell(\lambda|X)$ or minimize $-\ell(\lambda|X)$ w.r.t. $\lambda$ using e.g. gradient descent or simulated annealing or some other algorithm for finding minima. We can restrict search space by imposing restrictions on the search space, e.g. finite unconditional mean $\phi_1 < 1$ or positive unconditional mean: $\phi_1 > 0$ & $\phi_0 > 0$. This increases the speed of convergence.

Adrian Beer
1337304

We can obtain the time series of return variance of apple via $\{\sigma_t^2\}_{t \in \{2,...,T\}} = \{\varepsilon_t^2\}_{t \in \{2,...,T\}}$ since

$$\mathbb{E}[\varepsilon_t] = 0, \quad \mathbb{V}(\varepsilon_t) = \mathbb{E}[\varepsilon_t^2] - \mathbb{E}[\varepsilon_t]^2 = \mathbb{E}[\varepsilon_t^2].$$

If our model is a good fit, then $\hat{\sigma}_t$ and $\sigma_t$ should not deviate a lot. Furthermore we can check if we ~~decorrelated the~~ variance $\{\sigma_t^2 - \hat{\sigma}_t^2\}_{t \in \{2,...,T\}}$ still has any serial correlation with e.g. Ljung-Box test. If we modelled the heteroscedasticity correctly there should be none.

Adrian Beer
1087304

## 2.2. Task B.2

Let $\{x_t\}_{t \in \{1,..,T\}} = \begin{bmatrix} x_1 \\ \vdots \\ x_T \end{bmatrix}$ be the returns. When it

follows a Gaussian ARMA(1,1) – GARCH(1,1) model the DGP is

$$x_t = \hat{x}_t + \varepsilon_t \qquad \text{where}$$

$$\hat{x}_t = \phi_0 + \phi_1 x_{t-1} + \theta_1 \varepsilon_{t-1} = E[x_t \mid \mathcal{F}_{t-1}] \qquad \text{and}$$

$$\varepsilon_t = \hat{\sigma}_t \eta_t \qquad \text{where} \quad \eta \sim \mathcal{N}(0,1) \quad \text{and}$$

$$\hat{\sigma}_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \hat{\sigma}_{t-1}^2 = V(\varepsilon_t \mid \mathcal{F}_{t-1}).$$

$\varepsilon_{t-1}^2$ is the realized variance at time $t-1$: $\sigma_{t-1}^2$.

It follows that $x_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(\hat{x}_t, \hat{\sigma}_t^2)$ so the joint

log-likelihood can be written as

$$\overbrace{(\phi_0, \phi_1, \theta_1, \alpha_0, \alpha_1, \beta_1, \hat{\sigma}_1, \varepsilon_1)}$$

— will not use $x_1$, because we would have to estimate $x_0$ and $\varepsilon_0$, etc. too many additional parameters

$$L(\lambda \mid X) = \sum_{t=3}^{T} \log f(x_t \mid \lambda, \mathcal{F}_{t-1}) + \log f(x_1 \mid \lambda) + f(x_2 \mid \lambda).$$

where $f(x_t \mid \lambda, \mathcal{F}_{t-1}) = \dfrac{1}{\sqrt{2\pi \hat{\sigma}_t^2}} \cdot e^{-\frac{(x_t - \hat{x}_t)^2}{2\hat{\sigma}_t^2}}$ is the probability

density function of $x_t$. The first two observations are

special because no conditional variance estimation & cond. mean

can be calculated for those observations, because we lack

the preceding data necessary.* We can either include

$\hat{\sigma}_1^2, \varepsilon_1$ in the parameter vector $\lambda$, so that they will be

estimated or use the unconditional values,

\* namely $\varepsilon_1$ and $\hat{\sigma}_1^2$

i.e. $\hat{\sigma}_1^2 = V(\hat{\sigma}_\varepsilon^2) = \dfrac{\alpha_0}{1 - \alpha_1 - \beta_1}$ and $E[\varepsilon] = 0 = \varepsilon_1$. This way we can "salvage" $x_2$ and use it for the MLE.

We can ignore and not use $f(x_1)$ and $f(x_2)$

in the $L(\lambda)$ function at all, if we have enough data, however we need

estimates for $\varepsilon_1$ and $\hat{\sigma}_2^2$, because of their recursive nature/definition.

Adrian Beer
1007304

Conditional return density 2-step ahead. For this
we need $E[x_{t+2} | \mathcal{F}_t]$ and $V(x_{t+2} | \mathcal{F}_t)$.

First the conditional mean:

$$E[x_{t+2} | \mathcal{F}_t] = E[E[x_{t+2} | \mathcal{F}_{t+1}] | \mathcal{F}_t] \quad \text{(law of iterated expectations)}$$

$$= E[\phi_0 + \phi_1 x_{t+1} + \theta_1 \varepsilon_{t+1} + \varepsilon_{t+2} | \mathcal{F}_t]$$

$$\underbrace{E[\varepsilon_{t+2} | \mathcal{F}_t] = 0}$$

$$= E[\phi_0 + \phi_1(\phi_0 + \phi_1 x_t + \theta \varepsilon_t + \varepsilon_{t+1}) + \theta_1 \varepsilon_{t+1} + \varepsilon_{t+2} | \mathcal{F}_t]$$

$$= \phi_0 + \phi_1 \phi_0 + \phi_1^2 x_t + \theta \varepsilon_t + 0 + \theta_1 \cdot 0 + 0 = \phi_0(1+\phi_1) + \phi_1^2 x_t + \theta \varepsilon_t$$

because $x_t$ and $\varepsilon_t$ are known in $\mathcal{F}_t$ and $E[\varepsilon_{t+h} | \mathcal{F}_t] = 0 \; \forall h.$

~~Now the conditional variance:~~

$$\sout{V[x_{t+2} | \mathcal{F}_t] = E[(x_{t+2} - E[x_{t+2}])^2] \text{ with the law}}$$

~~of iterated expec. and the calculation from above we~~

~~arrive at~~ $\sout{V[x_{t+2} | \mathcal{F}_t] = E[((\theta_1 + \phi_1)\varepsilon_{t+1} + \varepsilon_{t+2})^2]}$

~~because the variance of everything known at time t is zero.~~

~~Since $\varepsilon_{t+1} \perp \varepsilon_{t+2}$~~

follows from above &
law of iterated expect. ✓

We model the conditional variance with GARCH, so

$$V_t(x_{t+2} | \mathcal{F}_t) = \sout{V(\varepsilon_{t+2} | \mathcal{F}_t)} E_t[(x_{t+2} - E[x_{t+2}])^2] = E_t[(\phi_1 \varepsilon_{t+1} + \theta \varepsilon_{t+1} + \varepsilon_{t+2})^2]$$

$$= E[\alpha_0 + \alpha_1 \varepsilon_{t+1}^2 + \beta_1 \hat{\sigma}_{t+1}^2 | \mathcal{F}_t] \quad \text{(GARCH)}$$

$$= E_t[\alpha_0 + \alpha_1 \varepsilon_{t+1}^2 + \beta_1(\alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \hat{\sigma}_t^2)]$$

$$= \alpha_0 + \underbrace{\alpha_1 \alpha_0 + \alpha_1^2 \varepsilon_t^2 + \alpha_1 \beta_1 \hat{\sigma}_t^2}_{} + \beta_1 \alpha_0 + \beta_1^2 \alpha_1 \varepsilon_t^2 + \beta_1^3 \hat{\sigma}_t^2$$

~~because $E[\varepsilon_{t+1}^2 | \mathcal{F}_t] =$~~ $V(\varepsilon_{t+1} | \mathcal{F}_t) = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \hat{\sigma}_t^2$ (GARCH)

$$= E_t[\phi_1^2 \varepsilon_{t+1}^2 + 2\phi_1 \theta_1 \varepsilon_{t+1}^2 + \theta_1^2 \varepsilon_{t+1}^2 + \varepsilon_{t+2}^2] \quad \text{because } \varepsilon_{t+1} \perp \varepsilon_{t+2}$$

$\boxed{\varepsilon_{t+1} \perp \varepsilon_{t+2}}$

$$E_t[E_{t+1}[\varepsilon_{t+2}^2]] = E_t[\alpha_0 + \alpha_1 \varepsilon_{t+1}^2 + \beta_1 \hat{\sigma}_{t+1}^2]$$

$$E_t[\hat{\sigma}_{t+1}^2] = E_t[\varepsilon_{t+1}^2]$$

Adrian Beer
1897304

$\Rightarrow \underset{t}{E}[\varepsilon_{t+2}^2] = \underset{t}{E}[\alpha_0 + (\alpha_1 + \beta_1)\varepsilon_{t+1}^2]$

Since $\underset{t}{E}[\varepsilon_{t+1}^2] = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \hat{\sigma}_t^2$ the equation

becomes:

$$\underset{t}{V}(x_{t+2}) = \phi_1^2(\alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \hat{\sigma}_t^2) + 2\phi_1\theta_1(\alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \hat{\sigma}_t^2)$$
$$+ \theta^2(\alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \hat{\sigma}_t^2) + \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \hat{\sigma}_t^2$$

$$\Rightarrow (\phi_1^2 + 2\phi_1\theta_1 + \theta^2 + 1)(\alpha_0 + \alpha_1 \varepsilon_t^2 + \hat{\sigma}_t^2)$$

$\overset{\Sigma=}{\phantom{x}}$ $\overset{\mu=}{\phantom{x}}$

Having determined $\underset{t}{V}(x_{t+2})$ and $\underset{t}{E}(x_{t+2})$ we can

write the conditional norm density (2-step ahead) as

$$f(x_{t+2} | \mathcal{F}_t) = \frac{1}{\sqrt{2\pi\Sigma}} \cdot e^{-\frac{(x_{t+2} - \mu)^2}{2\Sigma}}.$$

Note: In the calculation of the conditional variance

we can neglect $\eta$, because its variance is always

assumed to be 1.