Estimating Semiparametric ARCH (∞) models by Kernel Smoothing methods

GARCH(1,1) model:

$$\sigma_t^2 = \beta \sigma_{t-1}^2 + \alpha + \gamma y_{t-1}^2 \qquad (1)$$

Motivation: To increase the flexibility of the class of models we use and to learn from this the shape of the volatility function without restricting it a prior to have certain shapes.

Engle and Ng(1993) propose PNP model $\sigma_t^2 = \beta \sigma_{t-1}^2 + m(y_{t-j})$, where m is a smooth but unknown function. And the issue we solve is how to estimate the function $m(\cdot)$ by kernel methods.

We define the volatility process model

$$\sigma_t^2(\theta, m) = \mu_t + \sum_{j=1}^{\infty} \psi_j(\theta) m(y_{t-j}) \qquad (2)$$

where $\mu_t \in \mathbb{R}$, $\theta \in \Theta \subset \mathbb{R}^p$, and $m \in \mathcal{M}$, where $\mathcal{M} = \{m: \text{measure}\}$. The coefficients $\psi_j(\theta)$ satisfy $\psi_j(\theta) \geq 0$ and $\sum_{j=1}^{\infty} \psi_j(\theta) < \infty$ for all $\theta \in \Theta$. The true parameter is θ_0 and true function is $m_0(\cdot)$

Following Drost and Nijman(1993), we have strong, semi-strong and weak form ARCH(∞) process.

Strong: $\frac{y_t}{\sigma_t} = \varepsilon_t$ $\quad \varepsilon_t$ is iid with mean 0 and variance 1, where

$\sigma_t^2 = \sigma_t^2(\theta_0, m_0)$

Semi-strong: $E(y_t \mid \mathcal{F}_{t-1}) = 0$ and $E(y_t^2 \mid \mathcal{F}_{t-1}) \equiv \sigma_t^2$
where $\mathcal{F}_{t-1}$ is the sigma field generated by the entire past history of the y process

Weak: $P(y_t \mid \mathcal{F}_{t-1}) = 0$, $P(y_t^2 \mid \mathcal{F}_{t-1}) \equiv \sigma_t^2$
where $P(y_t \mid \mathcal{F}_{t-1})$ denotes the best linear predictor of y_t in terms of $y_{t-1}, \ldots$ $y_{t-1}^2, \cdots, 1$

we define θ_0, m_0 be defined as the minimizers of the population least squares criterion function

$$S(\theta, m) = E\left[\left\{ y_t^2 - \sum_{j=1}^{\infty} \psi_j(\theta) m(y_{t-j}) \right\}^2\right] \qquad (5)$$

and let $\sigma_t^2 = \sum_{j=1}^{\infty} \psi_j(\theta_0) m_0(y_{t-j})$. It is well defined only when $E(y_t^4) < \infty$

There is a special case that $\psi_j(\theta) = \theta^{j-1}$, with $0 < \theta < 1$, and we obtain

$$\sigma_t^2 = \theta \sigma_{t-1}^2 + m(y_{t-1}) \qquad (6)$$

$$(2) \Rightarrow \sigma_t^2 = \mu_t + \sum_{j=1}^{\infty} \theta^{j-1} m(y_t) \cdot L^j \quad \text{let } \mu_t = 0$$

$$\sigma_t^2 = \frac{m(y_{t-1})}{1 - \theta L}$$

$$\sigma_t^2 - \theta L \sigma_t^2 = m(y_{t-1}) \quad \sigma_t^2 = \theta \sigma_{t-1}^2 + m(y_{t-1})$$

When $m(y) = \alpha + \gamma y^2$, it is consistent with a stationary GARCH(1,1)

There ~~is a~~ are also other models as special cases including symmetric and asymmetric like $m(y) = \alpha + \gamma y^2 + \delta y^2 I(y<0)$, $m(y) = \alpha + \gamma(y+\delta)^2$, $m(y) = \alpha + \gamma |y|^\delta$ and we call $m(\cdot)$ the "new impact function". It determines the way in which the volatility is affected by shocks to y.

If m were known, it would be straightforward to estimate θ from some likelihood or least squares criterion.

Linear Characterization.

We define $\tilde{y}_t^2 = y_t^2 - \mu_t$.

$$E(\tilde{y}_t^2 \mid y_{t-j} = y) = \psi_j(\theta_0) m(y) + \sum_{k \neq j}^{\infty} \psi_k(\theta_0) E[m(y_{t-k}) \mid y_{t-j} = y]$$

Here we minimize creterion function (5) respect to m.

we write $m = m_0 + \epsilon f$ where m_0 is the the correct function.. we take differentiate respect to ϵ and setting $\epsilon = 0$ ~~ex~~ then obtain.

$$E\left[\left\{ y_t^2 - \sum_{j=1}^{\infty} \psi_j(\theta) m_0(y_{t-j}) \right\}\left\{ \sum_{l=1}^{\infty} \psi_l(\theta) f(y_{t-l}) \right\}\right] = 0$$

$$\Rightarrow \sum_{j=1}^{\infty} \psi_j(\theta) E[y_0^2 f(y_{-j})] - \sum_{j=1}^{\infty} \sum_{l=1, l \neq j}^{\infty} \psi_j(\theta) \psi_l(\theta) E[m_0(y_{-j}) f(y_{-l})] = \sum_{j=1}^{\infty} \psi_j^2(\theta) E[m_0(y_{-j}) f(y_{-j})]$$

$$E[y_0^2 f(y_{-j})] = E[y_0^2 \mid y_j = y] p_0(y)$$

We have

$$\sum_{j=1}^{\infty} \psi_j E[Y_0^2 \mid Y_{-j}=y] P_0(y) - \sum_{t=1}^{\infty}\left(\sum_{l=1}^{\infty} \psi_{t+l}(\theta)\psi_l(\theta)\right) E[m_0(Y_0) \mid Y_{j-l}=y]\cdot P_0(y)$$

$$= \sum_{j=1}^{\infty} \psi_j^2(\theta)\, E[m_0(Y_{-j}) f(Y_{-j})]$$

$$= \sum_{j=1}^{\infty} \psi_j^2(\theta)\, E[m_0(Y_{-j}) \mid Y_{-j}=y] P_0(y)$$

$$= \sum_{j=1}^{\infty} \psi_j^2(\theta)\, E[m_0(Y_0) \mid Y_0=y] P_0(y)$$

$\therefore$

$$\therefore \sum_{j=1}^{\infty} \psi_j E[Y_0^2 \mid Y_{-j}=y] P_0(y) - \sum_{t=1}^{\infty}\left(\sum_{l=1}^{\infty} \psi_{t+l}(\theta)\psi_l(\theta)\right) E[m_0(Y_0) \mid Y_{j-l}=y]\cdot P_0(y)$$

$$= \sum_{j=1}^{\infty} \psi_j^2(\theta) E[m_0(Y_0) \mid Y_0=y] P_0(y)$$

$$\sum_{j=1}^{\infty} \psi_j E[Y_0^2 \mid Y_{-j}=y] - \sum_{t=1}^{\infty}\left(\sum_{l=1}^{\infty} \psi_{t+l}\psi_l(\theta)\right) E[m_0(Y_0) \mid Y_{j-l}=y]$$

$$= \sum_{j=1}^{\infty} \psi_j^2(\theta)\, m_0(y)$$

divided by $\sum_{j=1}^{\infty} \psi_j^2(\theta)$

$$m_0(y) = \underbrace{\left[\frac{\sum_{j=1}^{\infty} \psi_j}{\sum_{j=1}^{\infty} \psi_j^2(\theta)}\right]}_{= \sum \psi_j^{+}(\theta)} \cdot E[Y_0^2 \mid Y_{-j}=y] - \underbrace{\left[\frac{\sum_{t=1}^{\infty}\left(\sum_{l=1}^{\infty} \psi_{l+t}(\theta)\psi_l(\theta)\right)}{\sum_{j=1}^{\infty} \psi_j^2(\theta)}\right]}_{= \sum \psi_t^{*}(\theta)} E[m_0(Y_0) \mid Y_0=y]$$

We have $m_\theta(y) = m_\theta^{*}(y) + \int H_\theta(y,x)\, m_\theta(x) P_0(x)\, dx$

where $\psi_j^{+}(\theta) = \psi_j(\theta) / \sum_{l=1}^{\infty} \psi_l^2(\theta)$ and $\psi_j^{*}(\theta) = \sum_{l=0} \psi_l \psi_{j+l}(\theta) / \sum_{l=1} \psi_l^2(\theta)$

and $H_\theta(y,x) = -\sum_{j=1}^{+\infty} \psi_j^{*}(\theta) \dfrac{P_{0,j}(y,x)}{P_0(y)P_0(x)}$

$$m_\theta^{*}(y) = \sum_{j=1}^{\infty} \psi_j^{+}(\theta)\, g_j(y)$$

We define M_c is the class of all bounded measurable functions that vanish outside $[-c, c]$

Then $m_{\theta,c}(y) = m_\theta^{*}(y) + \int_{-c}^{c} H_\theta(y,x)\, m_{\theta,c}(x) P_0(x)\, dx$

Simplify: $\Longrightarrow$ $m_\theta = m_\theta^{*} + H_\theta m_\theta$

where for each $\theta \in \Theta$, $\mathcal{H}_\theta$ is a self-adjoint linear operator on the Hilbert space of functions m that are defined on $[-c,c]$ with norm $\|m\|_2^2 = \int_{-c}^{c} m(x)^2 P_\theta(x)dx$

we have some ~~est~~ necessary assumptions for the next steps

Assumption 1: $\sup_{\theta \in \Theta} \int_{-c}^{c}\int_{-c}^{c} \mathcal{H}_\theta(x,y)^2 P_\theta(x) P_\theta(y) dx dy < \infty$

Assumption 2: is made to confirm $E[(\sum_{j=1}^{\infty} \psi_j(\theta) m(Y_{t-j}))^2] > 0$

Assumption 3: is a continuity condition.

using $E[(\sum_{j=1}^{\infty} \psi_j(\theta) m(Y_{t-j}))^2] > 0$ we obtain

$$\int m^2(x) P_\theta(x) dx - \int_{-c}^{c} m(x) \mathcal{H}_\theta m(x) P_\theta(x) dx$$

let λ to be the eigenvalue of $\mathcal{H}_\theta$, ~~the~~ and m are eigenfunctions.

we have $\int m^2(x) P_\theta(x) dx - \lambda \int m^2(x) P_\theta(x) dx > 0$

so $\lambda < 1 \Rightarrow I - \mathcal{H}_\theta$ has eigenvalues greater than 0. so positive definite. hence invertible and bounded by $(1-\gamma)^{-1}$

$$\|(1-\lambda)^{-1} m\|_2 = (1-\lambda)^{-1} \|m\|_2 \le (1-\gamma)^{-1}$$

so we can rewrite to obtain

$$m_\theta = (I - \mathcal{H}_\theta)^{-1} m_\theta^*$$

2.2 likelihood function.

Similar to Least squared linear criteria function., we have an alternative characterization of m_{θ}, θ in terms of Gaussian likelihood.

The Gaussian log likelihood function:

$$LLF = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\log \sigma_t^2 - \frac{1}{2}\sum_{i=1}^{N}\frac{\epsilon_t^2}{\sigma_t^2}$$

our criterion function: $l(\theta, m) = E[\log \sigma_t^2(\theta, m) + \frac{y_t^2}{\sigma_t^2(\theta, m)}]$

maximize LLF $\Longrightarrow$ minimize $l(\theta, m)$

$$\frac{\partial l(\theta,m)}{\partial(m)} = E\left[\frac{\Sigma \psi_j(\theta)}{\sigma_t^2(\theta,m)} + \frac{(-y_t^2)(\Sigma \psi_j(\theta))}{\sigma_t^4(\theta,m)}\right] \text{ (first-order condition)}$$

$$= \Sigma \psi_j(\theta) E\left[\sigma_t^{-4}(\theta,m)\{\sigma_t^2(\theta,m) - y_t^2\} \mid y_{t-j} = y\right] = 0$$

It is a non-linear equation respect to m.
So we use similar methods as Hastie and Tibshironi (Just Taylor Expansion)

$$\frac{\partial l}{\partial \eta} = \frac{\partial l}{\partial \eta_0} + \frac{d^2 l}{d\eta_0^2}(\eta - \eta_0)$$

So here we use similar method on l respect to m to linearize it, and obtain.

$\bar{m}_\theta = \bar{m}_\theta^* = \bar{\mathcal{H}}_\theta \bar{m}_\theta$, the definition of $\bar{m}_\theta^*$, $\bar{\mathcal{H}}_\theta$ are described in the paper P780.

The two methods based on different criterion function will bring different results ($\bar{m}_\theta$ and m_θ), ~~only~~ For strong / semi-strong form, we get $\bar{m}_{\theta_0} = m_{\theta_0}$

Estimation:

For different method (likelihood / least squared), we need different information for our estimation. The difference is that for $\bar{m}_\theta$, we need σ_t^2 to estimate it.

So the process will be

Estimate $\hat{m}_\theta^*$, $\hat{\mathcal{H}}_\theta$ $\rightarrow$ $\hat{m}_\theta$ $\rightarrow$ $\hat{\theta}$ $\rightarrow$ $\hat{\sigma}_t^2$ $\rightarrow$ $\hat{\bar{m}}_\theta^*$ and $\hat{\bar{\mathcal{H}}}_\theta$ $\rightarrow$ $\bar{m}_\theta$ $\rightarrow$ $\bar{\theta}$

While estimating m_θ^* and $\mathcal{H}_\theta$, we just use local linear (or other Kernel) methods to estimate $\mathbb{E}\, g_j(y) = E[\hat{Y}_t^2 \mid Y_{t-j} = y]$, $\hat{p}_{0,j}(y, x)$, $\hat{p}_0(y)$, $\hat{p}_0(x)$

Based on $\mathcal{H}_\theta$ and m_θ^*, we can estimate m_θ. Then we use the first order condition of Least squared criterion function to estimate θ. Then based on $\hat{m}_\theta$ and $\hat{\theta}$,

$$\hat{\sigma}_t^2 = \max\{\mu_t + \sum \psi_j(\theta)\, \hat{m}_\theta(Y_{t-j}), \epsilon\}$$

the may just confirm $\hat{\sigma}_t^2 > 0$

next, based on Kernel methods, we can obtain $\hat{g}_j^a$, $\hat{g}_j^b$, $\hat{g}_j^c$, hence $\hat{m}_\theta^*$ and $\hat{\mathcal{H}}_\theta$. Then use the first order condition to minimize $l(\theta, m)$ and obtain $\hat{\theta}$.

$\tilde{m}_\theta = \hat{m}_\theta^* + \hat{\mathcal{H}}_\theta \tilde{m}_\theta$