PART I: Modeling Real Exchange rates

Let the sample of real exchange rates$\{x_t\}_{t=1}^T$ be a subset of the realized path of a strictly stationary and ergodic (SE) time-series $\{x_t\}_{t\in\mathbb{Z}}$ with bounded moments of fourth order $\mathbb{E}|x_t|^4 < \infty$. Consider a Gaussian Exponential SESTAR model for the real exchange rate:

$$x_t = \alpha + g(x_{t-1};\boldsymbol{\theta})(x_{t-1}-\mu) + \epsilon_t \text{ where } \{\epsilon_t\}_{t\in\mathbb{Z}} \sim \text{NID}(0,\sigma_\epsilon^2), \tag{1}$$

$$g(x_{t-1};\boldsymbol{\theta}) := \delta + \frac{\gamma}{1+\exp(\beta(x_{t-1}-\mu)^2)} \text{ for every } t\in\mathbb{Z}. \tag{2}$$

Suppose that the parameters $\boldsymbol{\theta} = (\alpha,\delta,\gamma,\beta,\sigma_\epsilon^2)$ of the model are estimated by maximum likelihood (ML) on a compact parameter space Θ with $\beta \geq 0$ and $\sigma_\epsilon^2 > 0$.

Exercise 1. The log likelihood function allows us to find the parameter vector $\boldsymbol{\theta}$ for which we can optimize the joint probability distribution of the observed data. Note that the distribution of x_t conditional on the data $D_{t-1} = x_{t-1}, x_{t-2}, \ldots$ is Gaussian, i.e. $x_t|D_{t-1} \sim \mathcal{N}(\alpha + g(x_{t-1};\boldsymbol{\theta})(x_{t-1} - \mu), \sigma_\epsilon^2)$. Consequently, we have that the conditional density of x_t given the past data is:

$$f(x_t|D_{t-1};\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}}\exp\left[-\frac{(x_t-\alpha-g(x_{t-1};\boldsymbol{\theta})(x_{t-1}-\mu))^2}{2\sigma_\epsilon^2}\right]. \tag{3}$$

Since we can factorize the joint distribution of $x_t, x_{t-1}, \ldots$, we have that:

$$f(x_t, x_{t-1},\ldots;\boldsymbol{\theta}) = f(x_t;\boldsymbol{\theta})\cdot\prod_{t=2}^T f(x_t|x_{t-1};\boldsymbol{\theta}). \tag{4}$$

The conditional log likelihood function follows by taking the natural logarithm of the joint distribution of the data as given in (4) and dropping the marginal distribution $f(x_t;\boldsymbol{\theta})$. The latter is done as we don't want to make explicit assumptions about the marginal distribution of $f(x_t;\boldsymbol{\theta})$. We multiply the conditional likelihood function by $1/T$ to make the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ an M-estimator. Note that we can do this as multiplying by this constant does not change the maximisation problem. Consequently we obtain an expression for the criterion function, i.e. the log likelihood function:

$$\begin{aligned}
L_T(\boldsymbol{\theta}|\boldsymbol{X}_T) &= \frac{1}{T}\sum_{t=2}^T \log f(x_t|x_{t-1};\boldsymbol{\theta}) &(5)\\
&= \frac{1}{T}\sum_{t=2}^T -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma_\epsilon^2 - \frac{(x_t-\alpha-g(x_{t-1};\boldsymbol{\theta})(x_{t-1}-\mu))^2}{2\sigma_\epsilon^2} &(6)\\
&= \frac{1}{T}\sum_{t=2}^T C + \mathcal{L}(x_t, x_{t-1},\boldsymbol{\theta}), &(7)
\end{aligned}$$

where $C = -\frac{1}{2}\log 2\pi$ and $\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta}) = -\frac{1}{2}\log\sigma_\epsilon^2 - \frac{(x_t-\alpha-g(x_{t-1};\boldsymbol{\theta})(x_{t-1}-\mu))^2}{2\sigma_\epsilon^2}$.

Exercise 2. The criterion function $L_T(\boldsymbol{\theta},\boldsymbol{X}_T)$ converges uniformly in probability to the limit deterministic function $L_\infty(\boldsymbol{\theta})$, i.e.

$$\sup_{\theta\in\Theta}|L_T(\boldsymbol{\theta},\boldsymbol{X}_T) - L_\infty(\boldsymbol{\theta})| \xrightarrow{p} 0 \text{ as } T\to\infty, \tag{8}$$

if the criterion function $L_T(\boldsymbol{\theta},\boldsymbol{X}_T)$ converges pointwise to $L_\infty(\boldsymbol{\theta})$ on a compact Θ and the sequence $\{L_T(\boldsymbol{\theta},\boldsymbol{X}_T)\}_{T\in\mathbb{N}}$ is SE and stochastically equicontinuous. Note that by the linearity of expectation we can rewrite (8) as:

$$\sup_{\theta\in\Theta}|L_T(\boldsymbol{\theta},\boldsymbol{X}_T)-L_\infty(\boldsymbol{\theta})| = \sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=2}^{T} C+\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})-\mathbb{E}\left[C+\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})\right]\right| \tag{9}$$

$$= \sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=2}^{T} \mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})-\mathbb{E}\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})\right|. \tag{10}$$

This implies that if $\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})$ converges uniformly, also the criterion function $L_T(\boldsymbol{\theta},\boldsymbol{X}_T)$ converges uniformly to the limit criterion function $L_\infty(\boldsymbol{\theta})$. To show that $\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})$ converges pointwise, we use the Law of Large Numbers (LLN) for SE sequences. Namely, if $\mathbb{E}\,|\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})|<\infty$ and the sequence $\{\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})\}$ is SE, then it follows that

$$\left|\frac{1}{T}\sum_{t=2}^{T} \mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})-\mathbb{E}\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})\right| \xrightarrow{p} 0 \text{ as } T\to\infty \text{ for every } \theta\in\Theta. \tag{11}$$

In addition, to show that the sequence $\{\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})\}$ is stochastically equicontinuous we assume that the continuously differentiable function $\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})$ is well behaved of order 1 on the compact parameter space Θ. Note that $\{\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})\}$ is in fact continuously differentiable on the compact parameter space, since then it holds that $\sigma_\epsilon^2>0$. Namely, in that case the stochastic equicontinuity condition

$$\mathbb{E}\sup_{\theta\in\Theta}\left\|\frac{\partial\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|<\infty \tag{12}$$

is implied by the moment bound $\mathbb{E}\,|\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})|<\infty$ for some $\theta\in\Theta$.

By Krengel's theorem, $\{\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})\}$ SE since $\{x_t\}_{t\in\mathbb{Z}}$ is SE and $\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})$ is a continuous hence measureable function. Moreover, using that as $\{x_t\}_{t\in\mathbb{Z}}$ is SE, $\mathbb{E}|x_t|^4<\infty$, Θ is closed and bounded set, and that the product of bounded sets is also bounded, we have that $\mathbb{E}\,|\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})|<\infty$ since

$$\mathbb{E}\,|\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})| \tag{13}$$

$$= \mathbb{E}\left|-\frac{1}{2}\log\sigma_\epsilon^2-\frac{(x_t-\alpha-g(x_{t-1};\boldsymbol{\theta})(x_{t-1}-\mu))^2}{2\sigma_\epsilon^2}\right| \tag{14}$$

$$= \frac{1}{2}\log\sigma_\epsilon^2+\frac{1}{2\sigma_\epsilon^2}\mathbb{E}\,|x_t-\alpha-g(x_{t-1};\boldsymbol{\theta})(x_{t-1}-\mu)|^2 \quad \text{(linearity } \mathbb{E}) \tag{15}$$

$$\leq \frac{1}{2}\log\sigma_\epsilon^2+\frac{c_1}{2\sigma_\epsilon^2}\left[\mathbb{E}|x_t|^2+|\alpha|^2+\mathbb{E}\,|g(x_{t-1};\boldsymbol{\theta})(x_{t-1}-\mu)|^2\right] \quad (c_n\text{-inequality}) \tag{16}$$

$$\leq \frac{1}{2}\log\sigma_\epsilon^2+\frac{c_1}{2\sigma_\epsilon^2}\left[\mathbb{E}|x_t|^2+|\alpha|^2+\left(|\delta|+\frac{|\gamma|}{2}\right)^2\mathbb{E}\,|(x_{t-1}-\mu)|^2\right] \quad (g(x_{t-1};\boldsymbol{\theta})|\leq|\delta|+|\gamma|/2) \tag{17}$$

$$\leq \underbrace{\frac{1}{2}\log\sigma_\epsilon^2}_{<\infty}+\underbrace{\frac{c_1}{2\sigma_\epsilon^2}}_{<\infty}\left[\underbrace{\mathbb{E}|x_t|^2}_{<\infty}+\underbrace{|\alpha|^2}_{<\infty}+\underbrace{\left(|\delta|+\frac{|\gamma|}{2}\right)^2}_{<\infty}\left(c_2\underbrace{\mathbb{E}|x_{t-1}|^2}_{<\infty}+c_2\underbrace{|\mu|^2}_{<\infty}\right)\right] \quad (c_n\text{-inequality}) \tag{18}$$

$$< \infty. \tag{19}$$

Consequently, we obtain pointwise convergence for $\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})$. The stochastic equicontinuity of $\{\mathcal{L}(x_t,x_{t-1},\boldsymbol{\theta})\}$ follows easily since we have that it is an SE sequence and $\{L_T(\boldsymbol{\theta},\boldsymbol{X}_T)\}$ is a well behaved function of order one with a bounded first moment. In turn, if we assume that the compact set Θ is also convex, we conclude that the log likelihood function $L_T(\boldsymbol{\theta},\boldsymbol{X}_T)$ converges uniformly to a limit deterministic function $L_\infty(\boldsymbol{\theta}) = \mathbb{E}\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})$, since $\mathbb{E}$ integrates out the stochastic part.

Exercise 3. By the classical consistency theory for M-estimators, the ML estimator $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0 \in \Theta$ if:

1. The log likelihood criterion $L_T(\boldsymbol{\theta}, \boldsymbol{X}_T)$ converges in probability uniformly over Θ to the limit deterministic function $L_\infty(\boldsymbol{\theta})$ as $T \to \infty$;
2. The parameter $\boldsymbol{\theta}_0 \in \Theta$ is the identifiably unique maximizer of the limit criterion $L_\infty(\boldsymbol{\theta})$.

In exercise 2, we have showed that if we assume that Θ is compact and that $\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})$ is a well-behaved function of order 1, that the log likelihood criterion converges uniformly to the limit deterministic criterion function $L_\infty(\boldsymbol{\theta})$. If we further assume that $\boldsymbol{\theta}_0 \in \Theta$ is the unique maximizer of $L_\infty(\boldsymbol{\theta})$ on Θ, i.e.

$$\hat{\boldsymbol{\theta}}_T \in \arg\max_{\boldsymbol{\theta}\in\Theta} \frac{1}{T}\sum_{t=2}^{T} C + \mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}), \tag{20}$$

where C and $\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})$ are as defined in (7), it is also identifiably unique as $L_\infty(\boldsymbol{\theta})$ is continuous on a compact Θ with $\sigma_\epsilon^2 > 0$. Consequently, we have that the maximum likelihood estimator, or M-estimator, $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0 \in \Theta$, i.e. $\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$ as $T \to \infty$.

We know that $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0 \in \Theta$. Now assume that $\hat{\boldsymbol{\theta}}_T$ is consistent M-estimator for $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$ on a compact Θ, then by the classical asymptotic normality theorem it follows that $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \Omega\Sigma\Omega')$ as $T \to \infty$ where $\Omega = \left(\mathbb{E}\nabla^2\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}_0)\right)^{-1}$ if:

1. The scaled criterion derivative is asymptotically normal at $\boldsymbol{\theta}_0$, i.e.
$$\sqrt{T}\frac{1}{T}\sum_{t=2}^{T} \nabla\mathcal{L}\left(x_t, x_{t-1}, \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma) \text{ as } T \to \infty. \tag{21}$$
2. The second derivative of the criterion converges uniformly, i.e.
$$\sup_{\boldsymbol{\theta}\in\Theta} \left|\left|\frac{1}{T}\sum_{t=2}^{T} \nabla^2\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}) - \mathbb{E}\nabla^2\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right|\right| \xrightarrow{p} 0 \text{ as } T \to \infty. \tag{22}$$
3. The second derivative of the limit criterion function $\mathbb{E}\nabla^2\mathcal{L}\left(x_t, x_{t-1}, \boldsymbol{\theta}_0\right)$ is invertible.

Since $\{\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\}$ is SE and continuously differentiable of any order for $\sigma_\epsilon^2 > 0$, we know that $\{\nabla^n\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\}$ is also SE for any $n \in \mathbb{Z}^+$ by Krengel's theorem. Moreover, if we assume that that $\mathcal{L}$ is well-behaved of order two, i.e. $\mathcal{L} \in \text{WB}(2)$, and $\nabla\mathcal{L} \in \text{WB}(1)$, $\nabla^2\mathcal{L} \in \text{WB}(1)$, then it follows that the bounded moment

$$\mathbb{E}\left|\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right|^2 < \infty \text{ for some } \boldsymbol{\theta} \in \Theta \tag{23}$$

implies that (by the simple moments for asymptotic normality)

$$\mathbb{E}\left|\left|\nabla\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}_0)\right|\right|^2 < \infty, \tag{24}$$
$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\left|\nabla^2\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right|\right| < \infty, \tag{25}$$
$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\left|\nabla^3\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right|\right| < \infty. \tag{26}$$

Clearly, it holds true that $\mathbb{E}\left|\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right|^2 < \infty$ for some $\theta \in \Theta$, since $\{x_t\}_{t\in\mathbb{Z}}$ is SE, $\mathbb{E}|x_t|^4 < \infty$, and Θ is closed and bounded set:

$$
\begin{aligned}
&\mathbb{E}\left|\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right|^2 && (27)\\
&= \mathbb{E}\left|-\frac{1}{2}\log\sigma_\epsilon^2 - \frac{(x_t-\alpha-g(x_{t-1};\boldsymbol{\theta})(x_{t-1}-\mu))^2}{2\sigma_\epsilon^2}\right|^2 && (28)\\
&\leq c_1\left|\frac{1}{2}\log\sigma_\epsilon^2\right|^2 + c_1\left|\frac{1}{2\sigma_\epsilon^2}\right|^2 \mathbb{E}\left|x_t-\alpha-g(x_{t-1};\boldsymbol{\theta})(x_{t-1}-\mu)\right|^4 \quad {\scriptstyle (c_n\text{-inequality and linearity } \mathbb{E})} && (29)\\
&\leq c_1\left|\frac{1}{2}\log\sigma_\epsilon^2\right|^2 + c_1\left|\frac{1}{2\sigma_\epsilon^2}\right|^2 c_2\left[\mathbb{E}\left|x_t\right|^4 + |\alpha|^4 + \mathbb{E}\left|g(x_{t-1};\boldsymbol{\theta})(x_{t-1}-\mu)\right|^4\right] \quad {\scriptstyle (c_n\text{-inequality})} && (30)\\
&\leq c_1\left|\frac{1}{2}\log\sigma_\epsilon^2\right|^2 + c_1\left|\frac{1}{2\sigma_\epsilon^2}\right|^2 c_2\left[\mathbb{E}\left|x_t\right|^4 + |\alpha|^4 + \left(|\delta| + \frac{|\gamma|}{2}\right)^4 \mathbb{E}\left|x_{t-1}-\mu\right|^4\right] \quad {\scriptstyle \left(g(x_{t-1};\boldsymbol{\theta})| \leq |\delta| + \frac{|\gamma|}{2}\right)} && (31)\\
&\leq c_1\underbrace{\left|\frac{1}{2}\log\sigma_\epsilon^2\right|^2}_{<\infty} + c_1\underbrace{\left|\frac{1}{2\sigma_\epsilon^2}\right|^2}_{<\infty} c_2\left[\underbrace{\mathbb{E}\left|x_t\right|^4}_{<\infty} + \underbrace{|\alpha|^4}_{<\infty} + \left(\underbrace{|\delta|}_{<\infty} + \underbrace{\frac{|\gamma|}{2}}_{<\infty}\right)^4 c_3\left(\underbrace{\mathbb{E}|x_{t-1}|^4}_{<\infty} + \underbrace{|\mu|^4}_{<\infty}\right)\right] \quad {\scriptstyle (c_n\text{-inequality})} && (32)\\
&< \infty. && (33)
\end{aligned}
$$

Consequently, we can easily fulfil condition (1) of the classical asymptotic normality theorem. Namely if the model is well specified, then the sequence $\{\nabla\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}_0)\}$ is a martingale difference sequence (mds), i.e. $\mathbb{E}\left[\nabla\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}_0)|x_{t-1}, x_{t-2}, ...\right] = 0$. Hence, by Billingsley's CLT for SE sequences it satisfies a CLT at $\boldsymbol{\theta}_0$ since $\{\nabla\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}_0)\}$ is SE and $\mathbb{E}\left|\left|\nabla\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}_0)\right|\right|^2 < \infty$ (see (24)). If the model is mis-specified, it also satisfies a CLT since $\{\nabla\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}_0)\}$ is L^p-approximable by a mixingale as $\mathcal{L} \in \text{WB}(1)$ and $\mathbb{E}\left|\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right| < \infty$.

Moreover, we also fulfil condition (2) of the classical asymptotic normality theorem. That is, since $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\left|\nabla^2\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right|\right| < \infty$, see (25), implies that second order derivative of the criterion is bounded in expectation, i.e. $\mathbb{E}\left|\left|\nabla^2\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right|\right| < \infty \ \forall\ \boldsymbol{\theta} \in \Theta$. In combination with the fact that $\{\nabla^2\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\}$ is SE, we obtain by the LLN for SE sequences pointwise convergence in probability of the second order derivative criterion function, i.e.

$$\left|\left|\frac{1}{T}\sum_{t=2}^{T}\nabla^2\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}) - \mathbb{E}\nabla^2\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right|\right| \xrightarrow{p} 0 \text{ as } T \to \infty \text{ for every } \theta \in \Theta. \tag{34}$$

Furthermore since $\mathbb{E}\sup_{\theta\in\Theta}\left|\left|\nabla^3\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta})\right|\right| < \infty$ (see (26)), the sequence of the second order derivative of the criterion function is also stochastically equicontinuous. Hence, the second derivative of the criterion converges uniformly in probability.

Finally, we also fulfil condition (3) since we have assumed that $\boldsymbol{\theta}_0$ is the unique maximizer of $L_\infty(\boldsymbol{\theta})$ for the consistency of $\hat{\boldsymbol{\theta}}_T$. That implies that $\mathbb{E}\nabla^2\mathcal{L}\left(x_t, x_{t-1}, \boldsymbol{\theta}_0\right)$ is invertible.

Consequently, since all conditions of the classical asymptotic normality theorem are satisfied, we conclude that $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \Omega\Sigma\Omega')$ as $T \to \infty$ where $\Omega = \left(\mathbb{E}\nabla^2\mathcal{L}(x_t, x_{t-1}, \boldsymbol{\theta}_0\right)$. This implies that the approximate distribution of the maximum likelihood estimator is: $\hat{\boldsymbol{\theta}}_T \overset{app}{\sim} \mathcal{N}(\boldsymbol{\theta}_0, \Omega\Sigma\Omega'/T)$.

Exercise 4. Under the null-hypothesis $H_0 : \gamma = 0$ the exponential SESTAR model becomes $x_t = \alpha + \delta(x_{t-1} - \mu) + \epsilon_t = (\alpha - \delta\mu) - \delta x_{t-1} + \epsilon_t$ which is an AR(1). Consequently, testing this null-hypothesis is essentially testing for the non linear dynamics of the real exchange rate. That is because γ adds cyclical dependence to the model. As stated in exercise 3, we under that under the null-hypothesis $\hat{\gamma} \overset{app}{\sim} \mathcal{N}(\gamma_0, \hat{\sigma}_\gamma) \Leftrightarrow \hat{\gamma} \overset{app}{\sim} \mathcal{N}(0, 0.0657)$. The probability of obtaining an estimate $\hat{\gamma} = 0.724$ or something even more extreme under the null-hypothesis is close to zero. That is, the p-value for this hypothesis test is: $2 \cdot \left(1 - \Phi\left(\frac{0.724}{0.0657}\right)\right) \approx 0$. However, to be completely sure that there exist non linear dynamic behavior of the real exchange rate, we should also test the hypothesis $H_0 : \beta = 0$. Again, the p-value of this hypothesis test is approximately zero. This implies that there indeed exist non linear dynamic behavior. Namely, for large differences between x_{t-1} and μ the dependence of the model increases (i.e. mean reversion when far from equilibrium), while for small differences the model shows lower dependence (i.e. random walk behavior).

Exercise 5. If we assume that the model is well specified and that the true parameter $\boldsymbol{\theta}_0 = (\alpha_0, \delta_0, \gamma_0, \beta_0, \sigma^2_{\epsilon;0}) = (98.58, 0.629, 0.724, 0.063, 98.12, 0.997)$. Note that $\{x_t(\boldsymbol{\theta}, x_1)\}_{t\in\mathbb{N}}$ is a random sequence initialized at $t = 1$ with value $x_1 \in \chi \subseteq \mathbb{R}$ and generated by the Markov Dynamical System $x_{t+1} = \phi(x_t, \boldsymbol{\epsilon}_t, \boldsymbol{\theta})\ \forall\, t \in \mathbb{N}$, where ϕ is a differentiable function. Consequently, we can use the Power-n contraction theorem to show that $\{x_t(\boldsymbol{\theta}, x_1)\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique SE sequence $\{x_t\}_{t\in\mathbb{Z}}$, i.e. the real exchange rate time-series, satisfying $\mathbb{E}|x_t|^n$ for some $n \geq 0$, if the following conditions hold:

1. $\{\epsilon_t\}$ is an n_ϵ-variate exogenous iid SE sequence,
2. $\exists\, x_1 \in \chi : \mathbb{E}|\phi(x_1, \boldsymbol{\epsilon}_t)|^n < \infty$,
3. $\sup_{x\in\chi} \mathbb{E}|\partial\phi(x, \boldsymbol{\epsilon}_t)/\partial x|^n < 1$.

Clearly, condition (1) holds since $\{\epsilon_t\}$ is iid. Moreover, condition (2) also holds for any $n \geq 0$ since $\{\epsilon_t\}$ is Gaussian and hence has bounded moments of any order:

$$
\begin{aligned}
&\mathbb{E}|\phi(x_1, \boldsymbol{\epsilon}_t)|^n && (35)\\
&= \mathbb{E}\left|\alpha + \left(\delta + \frac{\gamma}{1+\exp(\beta(x_1-\mu)^2)}\right)(x_1-\mu) + \epsilon_t\right|^n && (36)\\
&= \mathbb{E}\left|98.58 + \left(0.629 + \frac{0.724}{1+\exp(0.063(x_1-98.12)^2)}\right)(x_1-98.12) + \epsilon_t\right|^n && (37)\\
&\leq c_1\left|98.58 + \left(0.629 + \frac{0.724}{1+\exp(0.063(x_1-98.12)^2)}\right)(x_1-98.12)\right|^n + c_1\mathbb{E}\left|\epsilon_t\right|^n \quad {\scriptstyle (c_n\text{-inequality})} && (38)\\
&< \infty. && (39)
\end{aligned}
$$

Lastly, we if we assume that

$$
\begin{aligned}
&\mathbb{E}\sup_{x\in\chi}\left|\frac{\partial\phi(x,\boldsymbol{\epsilon}_t)}{\partial x}\right|^n = && (40)\\
&\mathbb{E}\sup_{x\in\chi}\left|\frac{\partial}{\partial x}\left(98.58 + \underbrace{\left(0.629 + \frac{0.724}{1+\exp(0.063(x-98.12)^2)}\right)}_{g(x_{t-1};\boldsymbol{\theta_0})}(x-98.12) + \epsilon_t\right)\right|^n = && (41)\\
&\mathbb{E}\sup_{x\in\chi}\left|\frac{\partial}{\partial x}\left(98.58 + g(x_{t-1};\boldsymbol{\theta_0})(x-98.12) + \epsilon_t\right)\right|^n = && (42)\\
&\mathbb{E}\sup_{x\in\chi}\left|\left[\frac{\partial}{\partial x}g(x;\boldsymbol{\theta_0})\right](x-98.12) + g(x;\boldsymbol{\theta_0})\right|^n < 1 && (43)
\end{aligned}
$$

holds for $n = 1, 2, 3, 4$, we also satisfy condition (3) and hence conclude that the real exchange rate time-series is stationary. This implies that we could interpret the behavior of the real exchange rate by the exponential SESTAR as given in question 3. Here we have seen x_t for values of x_{t-1} close to its mean it shows non-stationary random walk behavior, while for values of x_{t-1} far from its mean, it shows mean reverting behaviour (i.e. stationarity). This result supports the purchase power parity, which says that real exchange rates should remain fixed over time under certain conditions. We would need the following for equation (43) to be true:

$$\mathbb{E}\sup_{x\in\mathcal{X}}\left|\left[\frac{\partial}{\partial x}g(x;\boldsymbol{\theta_0})\right](x-98.12)+g(x;\boldsymbol{\theta_0})\right|^n<1 \tag{44}$$

$$\Leftarrow \mathbb{E}\sup_{x\in\mathcal{X}}\left|\left[\frac{\partial}{\partial x}g(x;\boldsymbol{\theta_0})\right](x-98.12)+g(x;\boldsymbol{\theta_0})\right|<1 \tag{45}$$

$$\Leftarrow \sup_{x\in\mathcal{X}}\left|\left[\frac{\partial}{\partial x}g(x;\boldsymbol{\theta_0})\right](x-98.12)+g(x;\boldsymbol{\theta_0})\right| \tag{46}$$

$$\overset{\text{subadditivity of }\sup|\cdot|}{\leq} \sup_{x\in\mathcal{X}}\left|\left[\frac{\partial}{\partial x}g(x;\boldsymbol{\theta_0})\right](x-98.12)\right|+\sup_{x\in\mathcal{X}}|g(x;\boldsymbol{\theta_0})| \tag{47}$$

$$\overset{\sup_{x\in\mathcal{X}}|g(x;\boldsymbol{\theta_0})|=0.629+\frac{0.724}{2}}{=} \sup_{x\in\mathcal{X}}\left|\left[\frac{\partial}{\partial x}g(x;\boldsymbol{\theta_0})\right](x-98.12)\right|+0.629+\frac{0.724}{2} \tag{48}$$

$$= \sup_{x\in\mathcal{X}}\left|\frac{-0.724\cdot 2\cdot 0.063(x-98.12)\exp(0.063(x-98.12)^2)}{(1+\exp(0.063(x-98.12)^2))^2}(x-98.12)\right|+0.991 \tag{49}$$

$$\overset{\text{postive homogeneity of of }\sup|\cdot|}{\leq} |-0.724\cdot 2\cdot 0.063|\cdot\sup_{x\in\mathcal{X}}\left|\frac{\exp(0.063(x-98.12)^2)}{(1+\exp(0.063(x-98.12)^2))^2}(x-98.12)^2\right|+0.991<1 \tag{50}$$

From the above equations we see that we would need:

$$|-0.724\cdot 2\cdot 0.063|\cdot\sup_{x\in\mathcal{X}}\left|\frac{\exp(0.063(x-98.12)^2)}{(1+\exp(0.063(x-98.12)^2))^2}(x-98.12)^2\right|<(1-0.991)=0.009 \tag{51}$$

But we cannot show this because $\sup_{x\in\mathcal{X}}\left|\frac{\exp(g(x)}{(1+\exp(g(x))^2}(x-98.12)^2\right|<k<\infty$, where $g(x)=0.063(x-98.12)^2$. Therefore we cannot prove the stationarity and the bounded moments because neither the Power-n contraction theorem nor Bougerol's theorem give conclusive results. Because the conditions are merely sufficient and not necessary. What we can say is the following: we have seen in the previous question, the real exchange rate show stationary behaviour in the SESTAR model as they show mean reverting behaviour, especially when values of x_{t-1} are far from the mean (i.e. equilibrium). This observation supports the purchase power parity. Given that the process is mean-reverting, we can conclude that also higher-order moments should be present and hence that it is reasonable to assume that the fourth moment exists is bounded in expectation.

Exercise 6. In general, for comparing and selecting models we must take the number of parameters and the log likelihood in consideration. In order to compare the exponential SESTAR and the AR(1) model, which is nested by the exponential SESTAR, we can use the modified Akaike's Information Criterion (MAIC) which is defined as

$$MAIC=L_T^1(\boldsymbol{X_T},\hat{\boldsymbol{\theta}}_T^1)-L_T^2(\boldsymbol{X_T},\hat{\boldsymbol{\theta}}_T^2)-c(p-q)\left(T\log(\log(T))\right)^{\frac{1}{2}}, \tag{52}$$

where $p\geq q$, c a strictly positive scalar, and $L_T^1(\boldsymbol{X_T},\hat{\boldsymbol{\theta}}_T^1)$ and $L_T^2(\boldsymbol{X_T},\hat{\boldsymbol{\theta}}_T^2)$ are the log likelihoods of model 1 and 2 respectively. In our case, we have that $T=198$ and use $c=0.1$ as this value is acceptable as a rule-of-thumb. Consequently, the $MAIC\approx 19.20$. This implies that the SESTAR is preferred over the AR(1) model. Also intuitively we could conclude that the exponential SESTAR is preferred over the AR(1) model. Namely, in exercise 4 we have showed that the null-hypothesis $H_0:\gamma=0$ is rejected at any reasonable confidence level. Consequently, we already concluded that the non-linear dynamic behavior of the SESTAR is relevant to describe the time-series. Note however that this type of argumentation is subject to an approximate normal distribution and which in turn may be inaccurate in finite samples.

Exercise 7. Under mis-speficication, the ML estimator $\hat{\boldsymbol{\theta}}_T$ of the exponential SESTAR and the AR(1) model are consistent for the pseudo-true parameter $\boldsymbol{\theta}_0 \in \Theta$, which is still the unique maximizer of the limit criterion function $L_\infty(\boldsymbol{\theta})$. Consequently, this still implies that $\boldsymbol{\theta}_0$ is such that it maximizes the likelihood of observing the data. In the case of mis-specification, the probability model does not contain the actual DGP. However, the parameters $\boldsymbol{\theta}_0$ in both models minimize the Kullback-Leibler distance, which implies that the distance of the probability model to the real DGP is still minimized. Hence, $\boldsymbol{\theta}_0$ still provides the best possible approximation for the true DGP in both models.

Exercise 8. In table 1 we have displayed the conditional expectations of X_{T+n} for $n = 1, 2, ..., 12$. The conditional expectations have been found by Monte Carlo simulations (10,000). Note that the MATLAB code can be found in the appendix.

Exercise 9. In table 1 we have displayed the lower- and upper bounds of the 90% confidence interval of the forecasts. Furthermore, we have also calculated the probability the real exchange rate will rise above 105.33 in the following month using 10,000 Monte Carlo simulations. In particular, the AR(1) model gives a probability of 0.4692, while the exponential SESTAR model yields a probability 0.0237. The fact that the exponential SESTAR model expects the exchange rate to revert back to its equilibrium can be seen from the forecasts of the conditional expectation. This effect is not seen by the AR(1) model. Note that the MATLAB code can be found in the appendix.

Exercise 10. If we compare the forecasts of the exponential SESTAR and the AR(1) model to the actual observed values, we immediately observe that the AR(1) fails to describe the dynamics of the exchange rate well. Namely, over a period of 12 months the exchange rate drops back to its mean (equilibrium value). This effect is in fact captured by the exponential SESTAR model. In figure 1 it is also clear that the conditional expectations of the exponential SESTAR model better describe the actual observed time-series. Logically, this is especially the case for the smaller forecast horizons. Note that the MATLAB code can be found in the appendix.

Exercise 11. In figure 2 we observe in the 24-month impulse response functions (IRFs) that the exponential SESTAR model reacts to a by returning back to the equilibrium rather quickly. On the other hand, the AR(1) model does not have this mean reverting behavior but remains rather stable after the shock. Note that the MATLAB code can be found in the appendix.

Exercise 12. In figure 3 we have displayed the time-series of the real-exchange rates, i.e. the *Euro15 vs Swiss Franc* (above) and the *Euro15 vs UK Pound Sterling*. For each time-series, we have estimated a Gaussian AR(1), AR(2), AR(3) and exponential SESTAR model by maximum likelihood. Note that before the estimation, we have performed simulations to perform whether the likelihood function converge to the parameters of interest. Note that the MATLAB code can be found in the appendix. In table 3 we have showed the estimation results for the exponential SESTAR. In table 2 we have showed the results for the AR(1), AR(2), and AR(3) models. Also the log likelihoods of each of the models can be found in table 2. In order to select one model over another, we will use the MAIC (see equation (52)), which is a likelihood based measure. In table 4 we have displayed the MAIC statistics for the comparison of the different models. It is clear that for both time-series the AR(3) is preferred in terms of MAIC. Note that is has also the highest likelihood. Based on these findings, we conclude that there is no evidence of SESTAR dynamics in the data.

PART II: Modeling Financial Returns

Let the sample of the S&P500 returns $\{x_t\}_{t=1}^T$ at our disposal be a subset of the realized path of a SE time-series $\{x_t\}_{t\in\mathbb{Z}}$ with bounded moments of eighth order $\mathbb{E}|x_t|^8 < \infty$. We consider modeling the returns using the asymmetric GARCH model:

$$x_t = \sigma_t \epsilon_t \ \forall t \in \mathbb{Z}, \text{with } \{\epsilon_t\}_{t\in\mathbb{Z}} \sim NID(0,1) \text{ and the update equation} \tag{53a}$$

$$\sigma_t^2 = \omega + \alpha(x_{t-1} - \delta)^2 + \beta\sigma_{t-1}^2 \tag{53b}$$

We let $\boldsymbol{\theta} := (\omega, \alpha, \delta, \beta)$ the parameter vector in the parameter space Θ which we define to be compact. Let :

$$0 < a \leq \omega \leq b < \infty, \quad 0 < a \leq \alpha \leq b < \infty, \quad 0 < a \leq \beta < 1, \quad -\infty < c \leq \delta \leq d < \infty \tag{54}$$

Exercise 1. We will estimate these parameters by maximum likelihood and hence we will first define the criterion function we will maximize. Noting that the parameter space Θ is compact then if $Q_T(\cdot,\cdot) : \mathbb{R}^T \times \Theta \to \mathbb{R}$ is continuous on both Θ and $\mathbb{R}^T$ we have that the estimator $\hat{\boldsymbol{\theta}}_T : \mathcal{E} \to \mathbb{R}$ both exists (by Bolzano-Weierstrass) and is measurable (by Krengel's Theorem) and satisfies $\hat{\boldsymbol{\theta}}_T \in \arg\max_{\boldsymbol{\theta}\in\Theta} Q_T(\boldsymbol{X}_T, \boldsymbol{\theta})$. So let us define this $Q_T(\boldsymbol{X}_T, \boldsymbol{\theta})$ first. As previously done we will use the conditional likelihood:

$$f(x_t, x_{t-1}, \dots) = \prod_{t=2}^{T} f(x_t|\mathcal{F}_{t-1}; \boldsymbol{\theta}). \tag{55}$$

We note that given the past information $\mathcal{F}_{t-1}$ we have that $\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})$ is known, $\mathbb{E}[x_t|\mathcal{F}_{t-1}] = 0$ and $\mathbb{V}ar[x_t|\mathcal{F}_{t-1}] = \sigma_t^2(\sigma_1^2, \boldsymbol{\theta})$ so that we get $x_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}))$. Then we have that $f(x_t|\mathcal{F}_{t-1}) = (2\pi\sigma_t^2(\sigma_1^2, \boldsymbol{\theta}))^{-\frac{1}{2}} \exp(-\frac{x_t^2}{2\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})})$. If we fill this in equation 54 and take the log of this likelihood we get that

$$L_T(\boldsymbol{X}_T, \boldsymbol{\theta}) = \sum_{t=2}^{T} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})) - \frac{1}{2}\frac{x_t^2}{\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})} \tag{56}$$

Now we note that given that the estimator exists and is measurable that the $\hat{\boldsymbol{\theta}}_T$ that maximizes the likelihood in equation 55 is the same $\hat{\boldsymbol{\theta}}_T$ that maximizes the following equation:

$$Q_T(\boldsymbol{X}_T, \boldsymbol{\theta}) = \frac{1}{T}\sum_{t=2}^{T} -\log(\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})) - \frac{x_t^2}{\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})} \tag{57}$$

The last statement is true because $\frac{1}{2}, \log(2\pi)$ and $\frac{1}{T}$ are all constants where we added the $\frac{1}{T}$ to get the criterion function $Q_T(\boldsymbol{X}_T, \boldsymbol{\theta})$ in M-estimator form so we can apply the theorems for M-estimators

We further define:

$$\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta}) := -\log(\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})) - \frac{x_t^2}{\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})} \tag{58}$$

Exercise 2. At some point we have to explore the consistency of this estimator $\hat{\boldsymbol{\theta}}_T \in \arg\max_{\boldsymbol{\theta}\in\Theta} Q_T(\boldsymbol{X}_T, \boldsymbol{\theta})$, where we let $\boldsymbol{\theta}_0$ be the parameter of interest and $Q_\infty(\boldsymbol{\theta}) : \Theta \to \mathbb{R}$ be the limit function of the criterion function $Q_T(\boldsymbol{X}_T, \cdot) : \Theta \to \mathbb{R}$. We also make the assumption that $Q_\infty(\boldsymbol{\theta}) < Q_\infty(\boldsymbol{\theta}_0)$ $\forall \boldsymbol{\theta} \in \Theta$ i.e that $\boldsymbol{\theta}_0$ is the unique maximizer of $Q_\infty(\boldsymbol{\theta})$. We will show $\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$ as $T \to \infty$ using the general consistency theorem in chapter 5. The first thing we need to show is the uniform convergence $\sup_{\theta\in\Theta} |Q_T(\boldsymbol{X}_T, \boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta})| \xrightarrow{p} 0$ as $T \to \infty$ using the Uniform Convergence Theorem for M-estimators in chapter 5, which we will do in this exercise and in exercise 3 we show the Consistency.

Proof. We note the following important points:

- Θ is compact and we assume it to be convex as well
- $\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta})$ is continuously differentiable on Θ of at least order 6 where given the parameter space we have defined in 54 we have that $\sigma_t^2(\sigma_1^2, \boldsymbol{\theta}) > 0$, and hence $Q_T(\boldsymbol{X}_T, \boldsymbol{\theta})$ is also continuously differentiable of at least order 6.

We now will show that:

$$1. \; \frac{1}{T}\sum_{t=2}^{T} \mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta}) \xrightarrow{p} Q_\infty(\boldsymbol{\theta}) \; \forall \, \boldsymbol{\theta} \in \Theta \text{ as } T \to \infty \text{(pointwise convergence)} \tag{59}$$

$$2. \; \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} || \frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta}) || < \infty \text{ (stochastic equicontinuity)} \tag{60}$$

1:

Proof. To prove this we need to apply the LLN for SE sequences, but we recognise that $\{\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta})\}$ can never be SE because it is initialised at t=1. Hence we will first show that $\{\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})\}_{t\in\mathbb{N}}$ converges to a limit SE sequence $\{\sigma_t^2(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ then use this show that $\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta})$ converges to an SE sequence $\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})$. We begin with the applying the power-n contraction theorem to $\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})$.

Proof. Let $\{\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})\}_{t\in\mathbb{N}}$ be initialized at t=1 with $\sigma_1^2 \in \mathbb{R}$ and we observe that $\sigma_t^2 := \phi(\sigma_{t-1}^2, x_{t-1}, \boldsymbol{\theta}) := \omega + \alpha(x_{t-1} - \delta)^2 + \beta\sigma_{t-1}^2 \; \forall t \in \mathbb{N}$ for some $\boldsymbol{\theta} \in \Theta$ is a Markov Dynamical System, that is differentiable on Θ.

1. First we note that $\{x_t\}$ is an SE sequence and it is functionally exogenous
2. Next we have that :

$$\begin{aligned}
\mathbb{E}|\phi(\sigma_1^2, x_{t-1})|^4 &= \mathbb{E}|\omega + \alpha(x_{t-1} - \delta)^2 + \beta\sigma_1^2|^4 && (61)\\
&\leq c\mathbb{E}|\omega + \alpha(x_{t-1} - \delta)^2|^4 + c\mathbb{E}|\beta\sigma_1^2|^4 \;\; {\scriptstyle (c_n\text{-inequality, c>0})} && (62)\\
&\leq cd\mathbb{E}|\omega|^4 + cd\mathbb{E}|\alpha(x_{t-1} - \delta)^2|^4 + c\mathbb{E}|\beta\sigma_1^2|^4 \;\; {\scriptstyle (c_n\text{-inequality, d>0})} && (63)\\
&= cd|\omega|^4 + cd|\alpha|^4\mathbb{E}|(x_{t-1} - \delta)|^8 + c|\beta|^4|\sigma_1^2|^4 \;\; {\scriptstyle (\text{because } \omega,\, \alpha,\, \beta \,\in\, \Theta \text{ constants, } \sigma_1^2 \,\in\, \mathbb{R})} && (64)\\
&\leq cd|\omega|^4 + cde|\alpha|^4\mathbb{E}|x_{t-1}|^8 + cde|\alpha|^4|\delta|^8 + c|\beta|^4|\sigma_1^2|^4 \;\; {\scriptstyle (c_n\text{-ineq., e>0},\delta \text{ const.})} && (65)\\
&< \infty && (66)
\end{aligned}$$

   It is clear that in equation 64 all the constants are finite because they reside in a compact space, and $\sigma_1^2 \in \mathbb{R}$ furthermore we assume that $\mathbb{E}|x_t|^8 < \infty$ and because $\{x_t\}$ is SE we have that $\mathbb{E}|x_{t-1}|^8 = \mathbb{E}|x_t|^8 < \infty$ which why we conclude 53.

3. Now we check the contraction condition:

$$\mathbb{E}\sup_{\sigma^2} |\frac{\partial}{\partial \sigma^2}\phi(\sigma^2, x_{t-1})|^4 = \mathbb{E}\sup_{\sigma^2} |\beta|^4 = |\beta|^4 < 1 \Leftarrow |\beta| < 1 \tag{67}$$

   We can then conclude by the power-n contraction theorem that $\mathbb{E}|\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})|^4 < \infty$, $\{\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})\}_{t\in\mathbb{N}}$ converges e.a.s. to an SE limit sequence $\{\sigma_t^2(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and $\mathbb{E}|\sigma_t^2(\boldsymbol{\theta})|^4 < \infty$.

□

Now that we have established that $\{\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})\}_{t\in\mathbb{N}}$ converges e.a.s. to an SE limit sequence $\{\sigma_t^2(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ we further note that $\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta})$ is continuous hence by the continuous mapping theorem we have that $\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta}) \xrightarrow{p} \mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})$ as $t \to \infty$. Furthermore the continuity of $\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})$

implies that it is measurable under the Borel $\sigma-$algebra hence we can apply Krengel's theorem to conclude that $\{\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})\}$ is an SE sequence. Now we check the moments of $\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})$. We first note that for the parameter space Θ we defined in **??** we have that $\sigma_t^2(\boldsymbol{\theta}) \geq a > 0$ this ensures that $\mathbb{E}|\sigma_t^2(\boldsymbol{\theta}))|^4 < \infty$ (from the power-n contraction theorem) implies that $\mathbb{E}|\log(\sigma_t^2(\boldsymbol{\theta}))|^4 < \infty$, moreover it implies that $c\mathbb{E}|\frac{x_t^2}{\sigma_t^2(\boldsymbol{\theta})}|^4 \leq \frac{c}{a^4}c\mathbb{E}|x_t|^8$ and $\mathbb{E}|x_t|^8 < \infty$ by assumption. These imply the following :

$$\mathbb{E}|\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})|^4 = \mathbb{E}| - \log(\sigma_t^2(\boldsymbol{\theta})) - \frac{x_t^2}{\sigma_t^2(\boldsymbol{\theta})}|^4 \tag{68}$$

$$\leq c\mathbb{E}|\log(\sigma_t^2(\boldsymbol{\theta}))|^4 + c\mathbb{E}|\frac{x_t^2}{\sigma_t^2(\boldsymbol{\theta})}|^4 < \infty \text{ (c_n-inequality, c>0, and the notes above)} \tag{69}$$

Clearly, having $\mathbb{E}|\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})|^4 < \infty$ implies that $\mathbb{E}|\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})| < \infty$ which allows us to finally apply the LLN for SE sequences to $\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})$ and conclude that $\frac{1}{T}\sum_{t=2}^T \mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta}) \xrightarrow{p} \mathbb{E}|\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})|\ \forall\ \boldsymbol{\theta} \in \Theta$ as $T \to \infty$ (pointwise convergence). Where $\mathbb{E}|\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})| = Q_\infty(\boldsymbol{\theta})$. ☐

2:

Proof. To prove the stochastic equicontinuity of $\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})$ we will make use of the well-behavedness of the function. We note the following:

1. Θ is compact
2. $\phi(\sigma_t^2, x_t, \boldsymbol{\theta}) := \sigma_{t+1}^2$ is continuously differentiable and well behaved of order of 2 in $\boldsymbol{\theta} \in \Theta$ and in $\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})$, and the power-n conditions are fulfilled as can be seen in equation (66) and (60)-(65), where n=4 which is more than enough (we need n=2).
3. $\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta})$, and $\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})$ are both continuously differentiable and well behaved of order 2 in both $\boldsymbol{\theta} \in \Theta$ and in $\sigma_t^2(\sigma_1^2, \boldsymbol{\theta})/\sigma_t^2(\boldsymbol{\theta})$, as shown in the previous section.
4. Also $\mathbb{E}|\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta})|^4 < \infty$, $\mathbb{E}|\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})|^4 < \infty$ from the power-n contraction theorem we applied in the previous section.

Then we can conclude that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} ||\frac{\partial}{\partial\boldsymbol{\theta}}\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta})|| < \infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} ||\frac{\partial}{\partial\boldsymbol{\theta}}\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta})|| < \infty$. ☐

Now that we have shown $\frac{1}{T}\sum_{t=2}^T \mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta}) \xrightarrow{p} \mathbb{E}(\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta}))\ \forall\ \boldsymbol{\theta} \in \Theta$ as $T \to \infty$ (pointwise convergence) $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} ||\frac{\partial}{\partial\boldsymbol{\theta}}\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})|| < \infty$ (stochastic equicontinuity), and is Θ compact and we assume it to be convex as well we have that:

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\sum_{t=2}^T \mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta}) - \mathbb{E}(\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta}))\right| \xrightarrow{p} 0 \text{ as } T \to \infty \tag{70}$$

☐

Exercise 3. We now finalise showing $\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$ as $T \to \infty$ using the general consistency theorem in chapter 5.

Proof. As stated before we have assumed that $Q_\infty(\boldsymbol{\theta}) < Q_\infty(\boldsymbol{\theta}_0)\ \forall\boldsymbol{\theta} \in \Theta$ i.e that $\boldsymbol{\theta}_0$ is the unique maximizer of $Q_\infty(\boldsymbol{\theta}) := \mathbb{E}(\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta}))$. Furthermore we have from question 2 that, $\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\sum_{t=2}^T \mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}), \boldsymbol{\theta}) - \mathbb{E}(\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta}))\right| \xrightarrow{p} 0$ as $T \to \infty$ where $\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta})$ is continuous and because it converges uniformly to $\mathbb{E}(\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta}))$ we have that $\mathbb{E}(\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta}))$ is continuous. Therefore because Θ is compact and assumed to be convex we can conclude that $\boldsymbol{\theta}_0$ is the identifiably unique maximizer of $\mathbb{E}(\mathcal{L}(x_t, \sigma_t^2(\boldsymbol{\theta}), \boldsymbol{\theta}))$ by the identifiably uniqueness theorem of chapter 5.

Because now we have that:

1. $\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\sum_{t=2}^{T}\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})-\ \mathbb{E}(\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta}))\right| \xrightarrow{p} 0$ as $T\to\infty$

2. $\boldsymbol{\theta}_0$ is the identifiably unique maximizer of $\mathbb{E}(\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta}))$.

We can conclude by the general consistency theorem of chapter 5 that $\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$ as $T\to\infty$. □

Now we prove the asymptotic normality of the estimator. We apply the asymptotic normality theorem for M-estimators of chapter 6. Besides the assumption that the parameter of interest $\boldsymbol{\theta}_0$ is the unique maximizer of $\mathbb{E}(\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta}))$ we further let $\boldsymbol{\theta}_0\in\text{int}(\Theta)$. We will show that $\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)\xrightarrow{d}\mathcal{N}(0,\Omega\Sigma\Omega^T)$ where $\Omega=(\mathbb{E}[\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)])^{-1}$ and Σ is the asymptotic variance matrix of $\sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)$.

Proof. We already established that $\hat{\boldsymbol{\theta}}_T\xrightarrow{p}\boldsymbol{\theta}_0$ as $T\to\infty$ and defined Θ as convex and compact. We need to prove the following things:

1. $\sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)\xrightarrow{d}\mathcal{N}(0,\Sigma)$.
2. $\sup_{\boldsymbol{\theta}\in\Theta}||\frac{1}{T}\sum_{t=2}^{T}\nabla^2\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})-\mathbb{E}[\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})]||\xrightarrow{p}0$ as $T\to\infty$
3. $\mathbb{E}[\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)]$ is invertible

If we have that

- **A**: $\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})\in\mathbb{C}^3$ and $\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})\in\text{WB}(4)$ in both $\boldsymbol{\theta}\in\Theta$ and $\sigma_{t+1}^2:=\phi(\sigma_t^2,x_t,\boldsymbol{\theta})$ with $\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})\in\text{WB}(2)$ and $\nabla^2\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})\in\text{WB}(2)$ in both $\boldsymbol{\theta}\in\Theta$
- **B**: $\phi(\sigma_t^2,x_t,\boldsymbol{\theta})\in\mathbb{C}^3$ and $\phi(\sigma_t^2,x_t,\boldsymbol{\theta})\in\text{WB}(4)$ with $\nabla\phi(\sigma_t^2,x_t,\boldsymbol{\theta})\in\text{WB}(2)$ and $\nabla^2\phi(\sigma_t^2,x_t,\boldsymbol{\theta})\in\text{WB}(2)$.
- **C**: $\mathbb{E}|\phi(\sigma_1^2,x_t)|^4<\infty$, and $\mathbb{E}\sup_{\sigma^2}|\frac{\partial}{\partial\sigma^2}\phi(\sigma^2,x_{t-1})|^4$ (already proven by the power-n contraction theorem).

Then by the theorem in chapter 6 we have that $\mathbb{E}|\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})|^4<\infty$ implies that $\mathbb{E}||\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)||^2<\infty$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}||\nabla^2\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})||<\infty$, and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}||\nabla^3\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})||<\infty$. For **A** we have that $\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})=-\log(\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}))-\frac{x_t^2}{\sigma_t^2(\sigma_1^2,\boldsymbol{\theta})}\in\mathbb{C}^n$ because $\sigma_t^2(\sigma_1^2,\boldsymbol{\theta})\geq a>0$ so certainly $\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})\in\mathbb{C}^3$. Moreover we assume the well-behavedness of the function and its derivatives. For **B** we have that $\phi(\sigma_t^2,x_t,\boldsymbol{\theta})=\omega+\alpha(x_t-\delta)^2+\beta\sigma_t^2\in\mathbb{C}^3$ for the parameter space defined in 54 and we also assume the well behavedness of this function and its derivatives.
We now prove 1:

Proof. In order to prove $\sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)\xrightarrow{d}\mathcal{N}(0,\Sigma)$ we note the following:

1. **A** implies $\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})$ is continuously differentiable and $\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})\in\text{WB}(1)$ in both $\boldsymbol{\theta}\in\Theta$ and $\phi(\cdot)$.

2. $\mathbb{E}||\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)||^2<\infty$ implies that $\mathbb{E}||\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)||<\infty$

3. $\{\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})\}$ for some $\boldsymbol{\theta}\in\Theta$ was already established to be continuously differentiable and SE, hence by krengel's theorem $\{\nabla\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)\}$ is also SE, and also well behaved by the previous arguments. This means that it is L^p-approximable by a mixingale sequence in the case of misspecification. If the model is well-specified then we would get for free that $\{\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})\}$ is a martingale sequence (uncorrelated with mean zero) and we can also apply the CLT there. Moreover the fact that $E||\nabla\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)||^2<\infty$ means we can apply the CLT for SE sequences to $\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)$.

Then we have that $\sqrt{T}\left[\frac{1}{T}\sum_{t=2}^{T}\nabla\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)-\mathbb{E}[\nabla\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)]\right]\xrightarrow{d}\mathcal{N}(0,\Sigma)$ as $T\to\infty$ by the CLT for SE sequences, where $\mathbb{E}[\nabla\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)]=0$ because $\boldsymbol{\theta}_0$ is the identifiably unique maximizer of $\mathbb{E}[\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})]$. Moreover because of points 1 and 2 listed above we also have that $\sqrt{T}\left[\frac{1}{T}\sum_{t=2}^{T}\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)-\mathbb{E}[\nabla\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)]\right]$
$=\sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\nabla\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)\xrightarrow{d}\mathcal{N}(0,\Sigma)$. (because the difference vanished to zero asymptotically due to the well behavedness properties).

□

Now we prove 2:

Proof. To prove $\sup_{\boldsymbol{\theta}\in\Theta}||\frac{1}{T}\sum_{t=2}^{T}\nabla^2\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})-\mathbb{E}[\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})]||\xrightarrow{p}0$ as $T\to\infty$ we note that:

1. $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}||\nabla^2\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})||<\infty$ implies $\mathbb{E}||\nabla^2\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})||<\infty\ \forall\boldsymbol{\theta}\in\Theta$.
2. $\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})$ is continuously differentiable of order 3 so $\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})$ is continuous hence measurable under the Borel σ-algebra and we can apply krengel's theorem to conclude that $\{\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})\}$ is also SE. Furthermore it follows from the well behavednes properties mentioned earlier that $\mathbb{E}||\nabla^2\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})||<\infty\ \forall\boldsymbol{\theta}\in\Theta$ implies $\mathbb{E}||\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})||<\infty\ \forall\boldsymbol{\theta}\in\Theta$.

Because of 1 and 2 we can now apply the LLN for SE sequence and conclude that $\frac{1}{T}\sum_{t=2}^{T}\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})\xrightarrow{p}\mathbb{E}[\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})]$ as $T\to\infty$ and hence by the lemma on the slides of the solutions of week 3 we have that $\frac{1}{T}\sum_{t=2}^{T}\nabla^2\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})\xrightarrow{p}\mathbb{E}[\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})]$.

Lastly we note that already established $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}||\nabla^3\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})||<\infty$ and we conclude that $\sup_{\boldsymbol{\theta}\in\Theta}||\frac{1}{T}\sum_{t=2}^{T}\nabla^2\mathcal{L}(x_t,\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}),\boldsymbol{\theta})-\mathbb{E}[\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})]||\xrightarrow{p}0$ as $T\to\infty$. □

For 3: all we need is to note that $\boldsymbol{\theta}_0$ is the unique maximizer of $\mathbb{E}[\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}),\boldsymbol{\theta})]$ by assumption and hence $\mathbb{E}[\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)]$ is invertible. And from 1, 2, 3 we have that by the asymptotic normality theorem of chapter 6 that $\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)\xrightarrow{d}\mathcal{N}(0,\Omega\Sigma\Omega^T)$ where $\Omega=(\mathbb{E}[\nabla^2\mathcal{L}(x_t,\sigma_t^2(\boldsymbol{\theta}_0),\boldsymbol{\theta}_0)])^{-1}$ □

Then finally to answer the question about the asymptotic distribution of $\hat{\boldsymbol{\theta}}_T$ we have that $\hat{\boldsymbol{\theta}}_T\overset{app}{\sim}\mathcal{N}(\boldsymbol{\theta}_0,\frac{1}{T}\Omega\Sigma\Omega^T)$.

Exercise 4. First we analyze the volatility update equation: $\sigma_t^2=\omega+\alpha(x_{t-1}-\delta)^2+\beta\sigma_{t-1}^2$ and in particular the effect of δ in this equation. The case where $\delta=0$ is trivial because then the model reverts back to the normal GARCH because it nest the GARCH model. Suppose $\delta<0$, then if the previous return $x_{t-1}>0$ then the term $(x_{t-1}-\delta)^2$ is amplified hence the volatility is amplified, but if the previous return $x_{t-1}<0$ then the term $(x_{t-1}-\delta)^2$ is in fact softened, hence when we have $\delta<0$ then large postive returns produce higher volatility than large negative returns. What we want to test is whether the converse is true i.e. whether or not large negative returns yield a higher volatility update. So suppose $\delta>0$ then when $x_{t-1}<0$ the term $(x_{t-1}-\delta)^2$ is amplified (recall $(\cdot)^2\geq 0$) and when $x_{t-1}>0$ the term $(x_{t-1}-\delta)^2$ is softened. So if we have evidence that $\delta_0>0$ then we can confirm the leverage effect.

From the previous section we have that $\hat{\boldsymbol{\theta}}_T\overset{app}{\sim}\mathcal{N}(\boldsymbol{\theta}_0,\frac{1}{T}\Omega\Sigma\Omega^T)$. The reported standards errors in the table presented in page 6 of the assignment are based on the estimates $\hat{\Omega}$ and $\hat{\Sigma}$, which may or may not be robust under misspecification, but when we answer this question we do this under the assumption that these estimated errors are robust under incorrect misspecification. In question 9 we discuss this issue further. Now let $\boldsymbol{\theta}_0:=[\omega_0,\alpha_0,\delta_0,\beta_0]$, under $H_0:\delta_0=0$ we have that $\hat{\delta}_T\overset{app}{\sim}\mathcal{N}(0,\hat{\sigma}_T^2)$, where $\hat{\sigma}_T^2$ is the estimated variance and $\sqrt{\hat{\sigma}_T^2}=0.0014$ as given in the table. Then the probability of obtaining $\hat{\delta}_T=0.013$ or something more extreme under H_0 is $2\cdot(1-\Phi(\frac{0.013}{0.0014}))\approx 2\cdot(1-\Phi(9.2857))\approx 0$ (two-sided p-value), which we certainly deem low enough for us to reject the null-hypothesis that $\delta_0=0$.

It is visible that the normal GARCH(1,1) is nested in this Asymmetric GARCH where we have $\boldsymbol{\theta}_{As-GARCH} \in \{[\omega, \alpha, \delta, \beta]^T\} \supseteq \{[\omega, \alpha, 0, \beta]^T\} \ni \boldsymbol{\theta}_{Normal-GARCH}$ for the parameter space Θ as we defined in 54. By rejecting the null it can be said that $\{[\omega, \alpha, 0, \beta]^T\}$ is rejected from the parameter space, thereby rejecting the normal GARCH probability model. We know then that given the found values a high probability that volatility of this data does in fact not follow a normal GARCH(1,1) model. But to test the leverage effect we in fact need to perform a one side test $H_0 : \delta_0 \leq 0$, where we consider getting the probability of obtaining $\hat{\delta}_T = 0.013$ or higher than that which is given by $(1 - \Phi(\frac{0.013}{0.0014})) \approx (1 - \Phi(9.2857)) \approx 0$ (one-sided) hence we can reject the null hypothesis for any reasonable confidence level.Hence there is evidence that for this dataset the leverage effect indeed exists, and this data set is the S&P500 so it is representative of financial returns.

Exercise 5. If we look closely at the log likelihoods and the AIC found in question (4) (the results of question (4) are deemed to be correct so we use these as our reference point) and question (5) we have that in (4) log likelihood is -3306.95 and AIC is 6621.9 whereas in (5) the log likelihood is -3196 and the AIC is 6798.4. So we have a higher log likelihood and a higher AIC. This should not be possible. If we look at the formula for AIC we see that $AIC = 2k - 2L_T(\boldsymbol{X}_T, \hat{\boldsymbol{\theta}})$ so we have $AIC_{Q4} = 8 + 2(3306.95) = 6621.9$ but $AIC_{Q5} = 6 + 2(3196.2) = 6398.4 \neq 6798.4$. So the AIC of 6798.4 is wrong it should be 6398.4.

Exercise 6. If we now consider the GARCH(1,1) model under $\boldsymbol{\theta}_0 := (\omega_0, \alpha_0, \beta_0) = (0.017, 0.095, 0.891)$ and assume the model to be well-specified (i.e. that the model is the DGP) we can check whether or not the data generated by:

$$x_t = \sigma_t \epsilon_t \qquad \{\epsilon_t\}_{t\in\mathbb{Z}} \sim NID(0,1) \tag{71}$$

$$\sigma_{t+1}^2 = \omega_0 + \alpha_0 \sigma_t^2 \epsilon_t^2 + \beta_0 \sigma_t^2 = \omega_0 + \sigma_t^2(\alpha_0 \epsilon_t^2 + \beta_0) \tag{72}$$

is stationary by applying Bougerol's Theorem. We claim that the data generated by this model (both volatility and returns) is in fact stationary.

Proof. We begin with the stationarity of the volatility. Let $\{\sigma_t^2(\sigma_1^2, \boldsymbol{\theta}_0)\}_{t\in\mathbb{N}}$ initialised at $t = 1$ be a sequence generated by the model with $\sigma_1^2 \in \mathbb{R}$. We note that $\phi(\sigma_t^2, \epsilon_t, \boldsymbol{\theta}_0) := \sigma_{t+1}^2 = \omega_0 + \sigma_t^2(\alpha_0 \epsilon_t^2 + \beta_0)$ is a differentiable Markov Dynamical for which we can check the Bougerol conditions:

$$1.\ \{\epsilon_t\} \text{ is SE and functionally exogenous} \tag{73}$$

$$2.\ \exists\ \sigma_1^2 \in \chi \subseteq \mathbb{R} \text{ s.t. } \mathbb{E}\log_+|\phi(\sigma_1^2, \epsilon_t, \boldsymbol{\theta}_0)| < \infty \tag{74}$$

$$3.\ \mathbb{E}\log\sup_{\sigma^2}\left|\frac{\partial}{\partial\sigma^2}\phi(\sigma^2, \epsilon_t, \boldsymbol{\theta}_0)\right| < 0 \tag{75}$$

For (1) we note that the sequence $\{\epsilon_t\}$ was generated prior to the volatility sequence and hence it is functionally exogenous furthermore because it is i.i.d. it follows that it is SE. For (2):

$$\mathbb{E}\log_+|\phi(\sigma_1^2, \epsilon_t, \boldsymbol{\theta}_0)| = \mathbb{E}\log_+[|\omega_0 + \alpha_0\sigma_1^2\epsilon_t^2 + \beta_0\sigma_1^2|] = \mathbb{E}\log_+[|0.017 + 0.095\sigma_1^2\epsilon_t^2 + 0.891\sigma_1^2|] \tag{76}$$

$$\leq \mathbb{E}\log_+[|0.017| + |0.095\sigma_1^2\epsilon_t^2| + |0.891\sigma_1^2|] \text{ (By the subadditivity of } |\cdot|) \tag{77}$$

$$= \mathbb{E}\log_+[0.017 + 0.095|\sigma_1^2||\epsilon_t^2| + 0.891|\sigma_1^2|] \text{ (By pos. homogeneity of } |\cdot|) \tag{78}$$

$$= \mathbb{E}\log_+[0.017 + 0.095|\sigma_1^2||\epsilon_t^2| + 0.891|\sigma_1^2|] < \infty \tag{79}$$

$$\Leftarrow \mathbb{E}[0.017 + 0.095|\sigma_1^2||\epsilon_t^2| + 0.891|\sigma_1^2|] = \underbrace{0.017 + 0.095|\sigma_1^2|\mathbb{E}|\epsilon_t^2| + 0.891|\sigma_1^2|}_{\text{by linearity of } \mathbb{E}|\cdot|} < \infty \tag{80}$$

For (3):

$$\mathbb{E}\log\sup_{\sigma^2}|\frac{\partial}{\partial\sigma^2}\phi(\sigma^2,\epsilon_t,\boldsymbol{\theta}_0)| = \mathbb{E}\log\sup_{\sigma^2}|\alpha_0\epsilon_t^2+\beta_0| \tag{81}$$

$$= \mathbb{E}\log|0.095\epsilon_t^2+0.891| < 0 \text{ (Because } \alpha_0 \text{ and } \beta_0 \text{ constants)} \tag{82}$$

$$\underbrace{\Leftarrow}_{\text{by Jensen's inequality}} \quad \mathbb{E}[|0.095\epsilon_t^2+0.891|] \leq \mathbb{E}[|0.095\epsilon_t^2|+|0.891|] \text{ (By the subadditivity of } |\cdot|) \tag{83}$$

$$= 0.095\cdot \underbrace{\mathbb{E}|\epsilon_t^2|}_{=1 \text{ for}\{\epsilon_t\}\sim NID(0,1)} +0.891 \text{ (By linearity of } \mathbb{E}|\cdot|) \tag{84}$$

From 1,2,3 we can conclude by Bougerol's theorem that $\{\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0)\}_{t\in\mathbb{N}}$ converges e.a.s. to a limit SE sequence $\{\sigma_t^2(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$. Now that we have established $\gamma^t|\sigma_t^2(\sigma_1^2,\boldsymbol{\theta}_0)-\sigma_t^2(\boldsymbol{\theta}_0)| \stackrel{a.s.}{\to} 0$ as $t\to\infty$ we can prove properties of the data $\{x_t\}$ generated by this DGP. We want to show $\gamma^t|x_t(\sigma_1,\boldsymbol{\theta}_0)-x_t(\boldsymbol{\theta}_0)| \stackrel{a.s.}{\to} 0$. First we note the following:

1. $f(x)=\sqrt{x}$ for $x\geq 0$ is a continuous function therefore f is measurable and hence by Krengel's Theorem we have that $\sqrt{\sigma_t^2(\boldsymbol{\theta}_0)}=\sigma_t(\boldsymbol{\theta}_0)$ is also SE
2. We have that $\mathbb{E}|\epsilon_t|^n<\infty$ for all $n>0$ because the innovations are Gaussian hence clearly the much weaker condition $\mathbb{E}\log_+|\epsilon_t|<\infty$ is also fulfilled. Moreover since the innovations are i.i.d., then we can apply the product e.a.s. convergence theorem to obtain that $\epsilon_t\gamma^t|\sigma_t(\sigma_1,\boldsymbol{\theta}_0)-\sigma_t(\boldsymbol{\theta}_0)| \stackrel{a.s.}{\to} 0$ as $t\to\infty$.

Now we have that

$$\gamma^t|x_t(\sigma_1,\boldsymbol{\theta}_0)-x_t(\boldsymbol{\theta}_0)| = \gamma^t|\epsilon_t\sigma_t(\sigma_1,\boldsymbol{\theta}_0)-\epsilon_t\sigma_t(\boldsymbol{\theta}_0)| \text{ } \sigma_t(\boldsymbol{\theta}_0) \text{ is SE by (1)} \tag{85}$$

$$= \epsilon_t\gamma^t|\sigma_t(\sigma_1,\boldsymbol{\theta}_0)-\sigma_t(\boldsymbol{\theta}_0)| \stackrel{a.s.}{\to} 0 \text{ as } t\to\infty \text{ by (2)} \tag{86}$$

Now we have shown that both the innovations and the data generated by this DGP is in fact stationary.

□

Exercise 7. As we have noted before the Asymmetric GARCH nests the normal GARCH. In mathematical terms $\mathbb{P}_\theta^{As-GARCH} := \{P_\theta^{As-GARCH} : \theta\in\Theta_{As-GARCH}\} \supseteq \{P_{\theta^*}^{GARCH} : \theta^*\in\Theta_{GARCH}\} := \mathbb{P}_{\theta^*}^{GARCH}$. Let P_0 be the true DGP. Then is we suppose that there exists a $\theta_0\in\Theta_{As-GARCH}$ such that $\mathbb{P}_\theta^{As-GARCH} \ni P_{\theta_0}^{As-GARCH} = P_0$ (i.e. Asymmetric GARCH is well specified). Then it is possible that the θ_0 is contained in both $\Theta_{As-GARCH}$ and in Θ_{GARCH} or in other terms $P_0\in \mathbb{P}_\theta^{As-GARCH}\bigcap\mathbb{P}_{\theta*}^{GARCH}$(then the normal GARCH would be well-specified as well) but it is also possible that θ_0 is contained only in $\Theta_{As-Garch}$ or in other terms: $P_0\in\underbrace{\mathbb{P}_\theta^{GARCH^C}}_{complement}\bigcup\mathbb{P}_\theta^{As-GARCH}$ (the the normal GARCH is misspecified). If on the other hand we suppose that the GARCH is well specified, there exists a $\theta_0\in\Theta^{GARCH}$ such that $P_{\theta_0}^{GARCH}=P_0$ then we can observe that $\mathbb{P}_{\theta*}^{GARCH}\subseteq\mathbb{P}_\theta^{As-GARCH}$ which by definition of a subset means that if $P_0=P_{\theta_0}^{GARCH}\in\mathbb{P}_{\theta*}^{GARCH}$ then $P_0\in\mathbb{P}_\theta^{As-GARCH}$. Hence if the GARCH is well specified than the Asymmetric GARCH is also well specified. Then clearly the contrapositive is then true, if the Asymmetric GARCH is misspecified then the GARCH is misspecified as well.

Exercise 8. When plotting the news impact curves we have used the following parameters $\boldsymbol{\theta} = (\omega,\alpha,\beta,\lambda)=(1,0.1,0.89,\lambda)$ where we vary the degrees of freedom λ to be either $2,5,10,50$. The results can be found in figure 4. It is visible from the plot that the Robust GARCH does what it is supposed to, because as the degrees of freedom increase the volatility produced by the update equation as a result of the previous information (news) becomes higher. And the volatility produced by the normal GARCH is the highest as a reaction to the news.

Exercise 9. In tables 5 and **??** the estimated parameters and the standard errors are given. Due to time constraints the following is to be noted: we have not implemented robust standard errors under misspecification, because we were unable to implement it correctly. The reported standard errors are valid under correct specification where we have used the following approximation under $H_0 : \boldsymbol{\theta}_0 = \boldsymbol{\theta}_*$ we have $\hat{\boldsymbol{\theta}}_T \sim \mathcal{N}(\boldsymbol{\theta}_*, \hat{\Sigma}_T^{-1} \cdot \frac{1}{T})$ where $\hat{\Sigma}_T^{-1} = (\frac{1}{T} \sum_{t=2}^{T} \nabla^2(\mathcal{L}(x_t, \sigma_t^2(\sigma_1^2, \boldsymbol{\theta}_*), \boldsymbol{\theta}_*)))^{-1}$. When looking at the estimated values we observe that the parameters found for the Yahoo, and Nokia are rather high, this could results in strong shocks in volatility as a results of shocks in returns especially which could affect the GARCH model.

Exercise 10. In figures 5 and **??** we have plotted the filtered volatilities of the GARCH and the Robust GARCH, it is immediately visible that the Normal GARCH is highly sensitive to outliers, whereas the Robust GARCH remains more stable.

Exercise 11. In table 7 we have displayed the Value-at-Risk (VaR) estimates implied by the filtered volatilities. The Value-at-Risk is defined as the maximum loss an investor can loose with a certain probability over a specified horizon. We have calculated the one-step-ahead VaR forecasts assuming that we have a long position in the stock. That is, the VaR is defined as

$$\text{VaR} = \hat{\sigma}_{t+1|t}\vartheta_{1-\alpha}, \tag{87}$$

where $\hat{\sigma}_{t+1|t}$ is the one-step ahead volatility forecast given all the data, $\vartheta_{1-\alpha}$ is the (1-α)th quantile of the cumulative return function of the log returns, and α is the confidence level of the VaR. Since in the normal GARCH model we essentially assume that returns are normally distributed, we have that $\vartheta_{1-\alpha} = \Phi^{-1}(1-\alpha)$. For the Robust GARCH we use a Student-t distribution. Consequently, we have that $\vartheta_{1-\alpha}$ is the standardized quantile of the Student-t distribution, i.e. $\vartheta_{1-\alpha} = q_{1-\alpha}(v)/\sqrt{v/(v-2)}$, where $q_{1-\alpha}$ is the quantile of the Student-t distribution with v degrees of freedom obtained thorugh the standard quantile function.

As mentioned in exercise 11, the normal GARCH produces higher implied volatility forecasts over the sample compared to the robust GARCH model. That is why we observe that the VaR forecasts for the normal GARCH are relatively high. Namely due to the higher volatility forecasts the investor has a higher VaR. Note that even though we have that the Robust GARCH uses a t-distribution, which implies that the tails of the return distribution are fatter and hence we can expect more extreme events to occur, we still have that the normal GARCH model due to its high volatility forecasts results in a higher VaR.

Exercise 12. To compare the GARCH model to the Robust GARCH model we use the modified Akaike's Information Criterion (MAIC). We have 3162 data points so $T = 3162$, and we let $c = 0.1$. For $L_T^1(\boldsymbol{X_T}, \hat{\boldsymbol{\theta}}_T^1)$ we choose the Robust GARCH (because, we want $p \geq q$, and the Robust GARCH has more parameters) and for $L_T^2(\boldsymbol{X_T}, \hat{\boldsymbol{\theta}}_T^2)$ we choose the GARCH. So a positive MAIC would constitute evidence in favour of the Robust GARCH and a negative MAIC the converse. We then have $p = 4, q = 3$ and we also calculate the penalty term $c(p-q)\left(T\log(\log(T))\right)^{\frac{1}{2}}$ exactly because we have $T = 3162$. To reiterate we use the formula of 52. The results are found in **??**. These results are very much in line in what we have found previously. For instance, if we look at the filtered volatility plots of Yahoo and Nokia we see that the for the GARCH the volatility is overestimated wheres the Robust GARCH remains stable. Hence, according to the plots and the MAIC statistics for these stocks, one should definitely choose the Robust GARCH. A negative MAIC in favor of the Normal GARCH is also plausible for moicrosoft. Choosing a Robust GARCH is also plausible for intel. However the volatility plots versus the NIC plots do not support the MAIC findings of choosing GE. Because when we look at the filtered volatility of GE we see that the volatility produced by the GARCH reacts very dramatically to extreme returns but the MAIC points to choosing the Normal GARCH. We know too little of the theoretical grounds of the MAIC to further investigate as to why this is possible.

Table 1: Conditional expecations and 90% CI

X_{T+n}	Exponential SESTAR			AR(1)		
	LB	**cond. $\mathbb{E}$**	**UB**	**LB**	**cond. $\mathbb{E}$**	**UB**
$n=0$	105.33	105.33	105.33	105.33	105.33	105.33
$n=1$	101.67	103.32	104.92	103.59	105.26	106.93
$n=2$	100.60	102.40	104.19	102.87	105.19	107.53
$n=3$	99.962	101.92	103.80	102.32	105.11	107.94
$n=4$	99.601	101.66	103.62	101.81	105.03	108.28
$n=5$	99.330	101.48	103.51	101.42	104.97	108.56
$n=6$	99.075	101.37	103.46	100.98	104.90	108.88
$n=7$	98.943	101.29	103.39	100.69	104.83	109.05
$n=8$	98.818	101.24	103.36	100.32	104.77	109.20
$n=9$	98.695	101.21	103.36	100.07	104.71	109.41
$n=10$	98.602	101.17	103.38	99.739	104.64	109.62
$n=11$	98.544	101.14	103.33	99.438	104.58	109.73
$n=12$	98.508	101.11	103.34	99.166	104.52	109.85

Note: this table shows the 1 to 12-step-ahead forecasts of the conditional expectation of the real exchange rate from the starting month April, 2009 for the exponential SESTAR and AR(1) model. In this month the real exchange rate was at a level of 105.33 (i.e. $n = 0$). Note that since the normality of the innovations was not rejected by a Jarque-Bera test, we have added 90% confidence bounds to the forecasts. The LB implies the lower bound of the confidence interval, UB is the upper bound, and the conditional expectation is given by cond. $\mathbb{E}$.

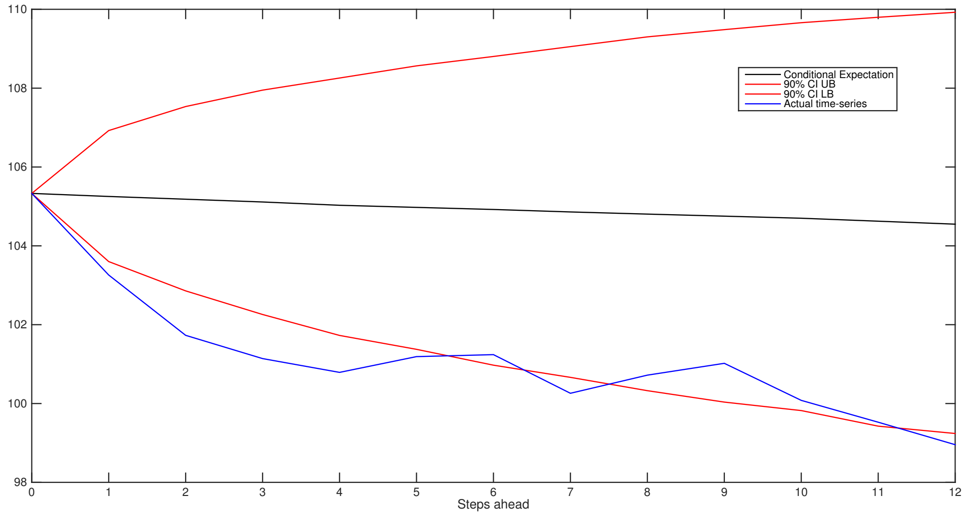
Table 2: Parameter estimates AR

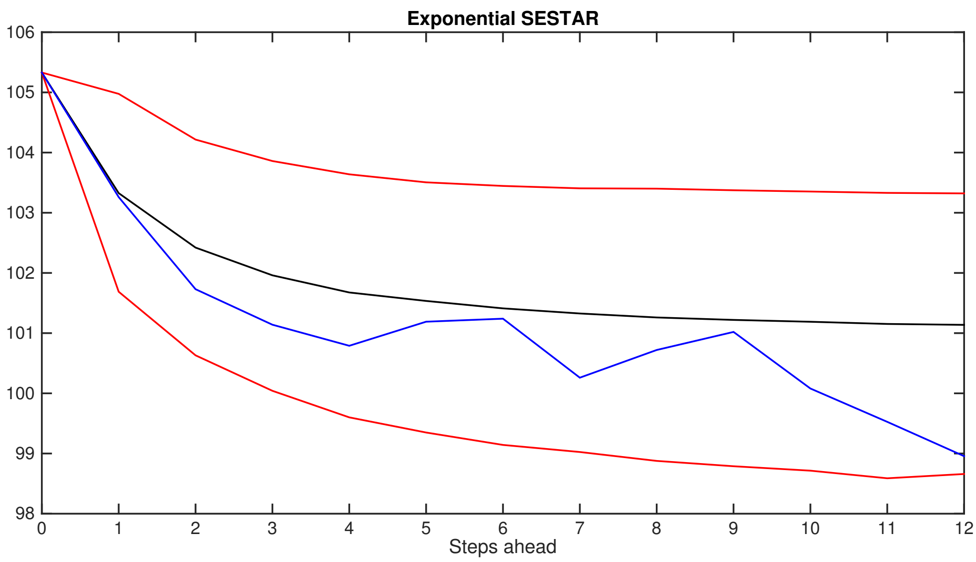
Euro15 vs UK Pound Sterling						
Parameter	**AR(1)**		**AR(2)**		**AR(3)**	
	Estimate	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
ϕ_1	0.97969	0.012364	1.076	0.061917	1.095	0.063053
ϕ_2			-0.089718	0.06187	-0.12014	0.0914
ϕ_3					0.011231	0.061935
ω	1.9217	1.1661	1.3126	1.1531	1.3191	1.1565
σ^2_ϵ	1.6405	0.14644	1.5771	0.14136	1.563	0.14001
LL	-418.279703		-411.760173		-409.2248	

Euro15 vs Swiss Franc						
Parameter	**AR(1)**		**AR(2)**		**AR(3)**	
	Estimate	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
ϕ_1	0.98027	0.01364	1.0841	0.063637	1.0883	0.063461
ϕ_2			-0.096381	0.063827	-0.11218	0.09327
ϕ_3					0.011428	0.063627
ω	2.0666	1.3906	1.297	1.4153	1.3177	1.4164
σ^2_ϵ	1.8729	0.16719	1.8982	0.17353	1.8642	0.16702
LL	-434.906304		-432.474945		-431.159678	

Note: this stable shows the maximum likelihood estimates and standard errors (SE) of the AR(1), AR(2), and AR(3) model, i.e. $x_t = \omega + \sum_{i=0}^{p} \phi_p x_{t-p} + \epsilon_t$ where p is the order of the AR model.

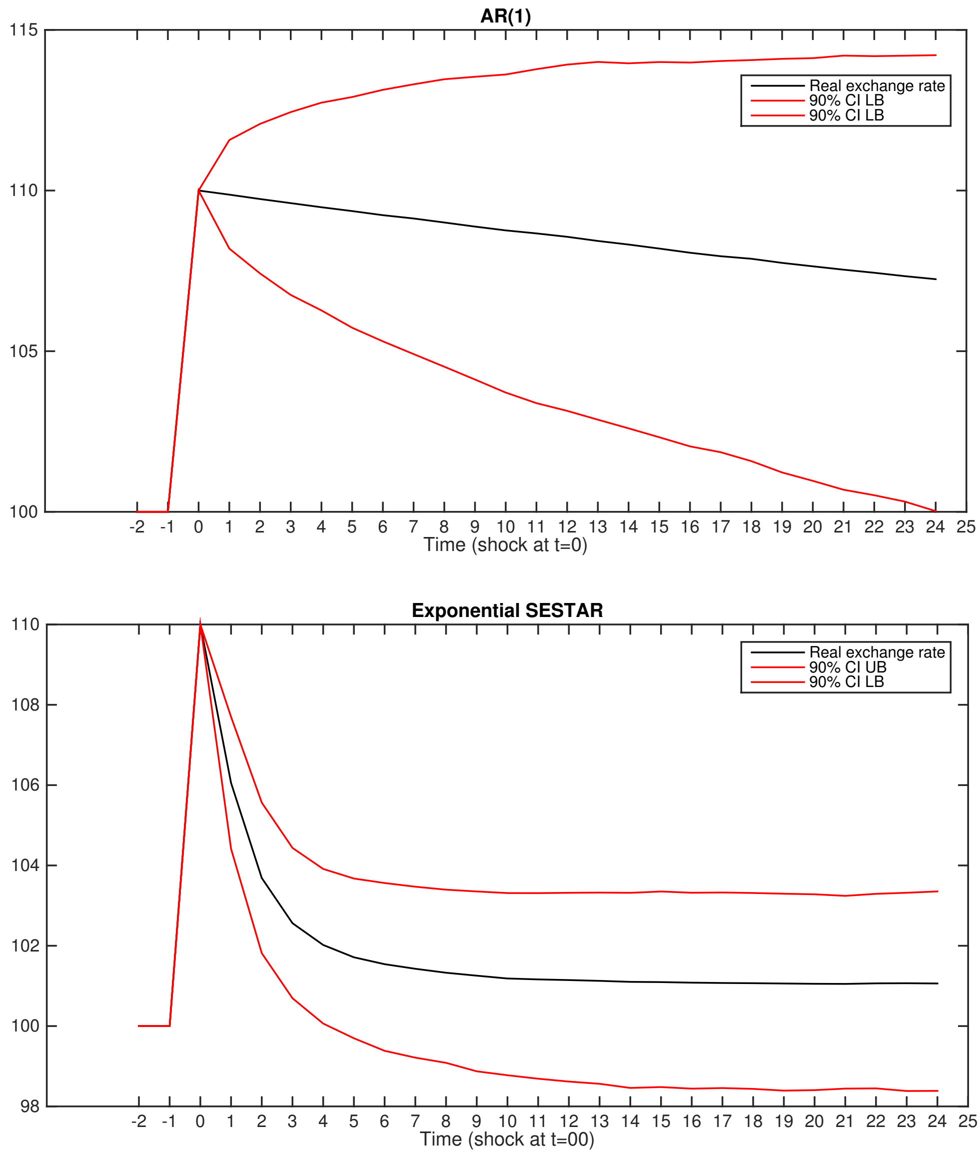
Figure 1: Conditional expecations and 90% CI





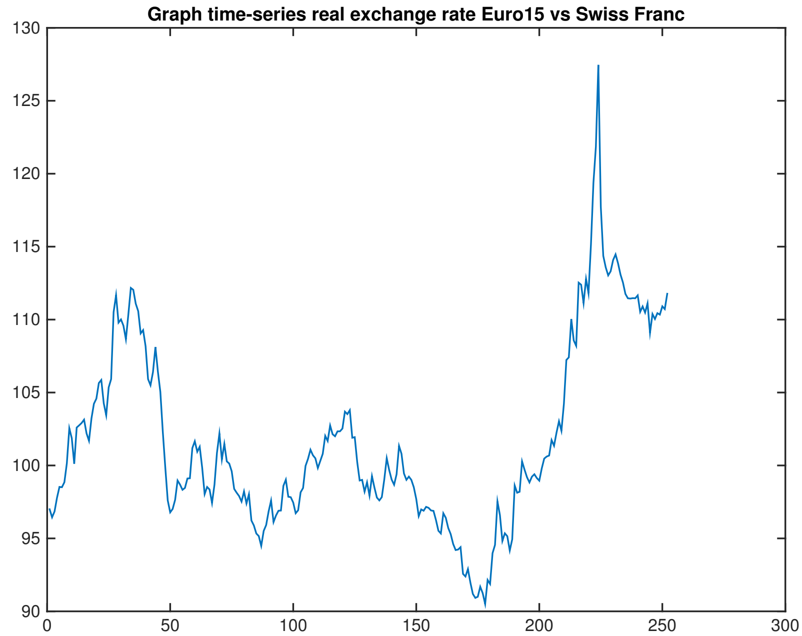
Note: this figure shows the plot of the conditional expectations and the 90% confidence bounds for the AR(1) model and the exponential SESTAR. For the sake of comparison, also the actual observed time-series is plotted.

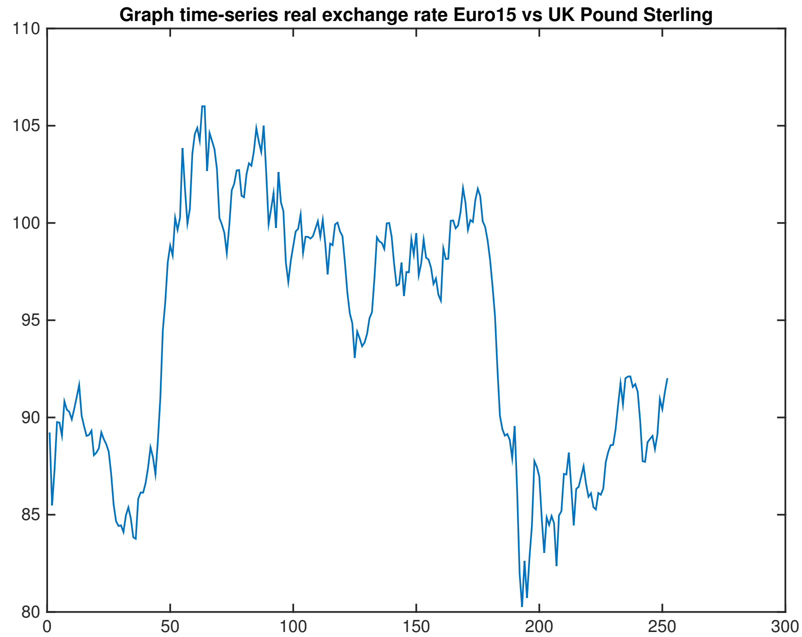
Figure 2: IRF and 90% CI



Note: this figure shows the 24-month impulse response functions (IRFs) using both the SESTAR and AR(1) model. The IRFs are generated by a positive shock of size 10, with origin at 100.

Figure 3: Time-series real exchange rates





Note: this figure shows the plot of the time-series real exchange rate of the *Euro15 vs Swiss Franc* (above) and the *Euro15 vs UK Pound Sterling* (below).

Table 3: Parameter estimates Exponential SESTAR

Parameter	Euro15 vs UK Sterling Pound		Euro15 vs Swiss Franc	
	Estimate	SE	Estimate	SE
α	97.958	0.06805	98.604	0.067659
μ	98.014	0.076664	98.481	0.06662
δ	0.981	0.013052	0.98042	0.014413
γ	-0.16836	0.50959	-0.045163	0.51133
β	0.48495	0.04217	0.60904	0.053249
σ_ϵ^2	1.6394	0.14891	1.8729	0.16535
LL	-418.225504		-434.902918	

Note: this stable shows the maximum likelihood estimates and standard errors (SE) of the Exponential SESTAR model (see equation (1)).

Table 4: MAIC statistics

	Euro15 vs Swiss Franc			Euro15 vs UK Sterling Pound		
	AR(1)	**AR(2)**	**AR(3)**	**AR(1)**	**AR(2)**	**AR(3)**
AR(2)	1.4522			5.5404		
AR(3)	1.7883	0.3361		7.0966	1.5562	
SESTAR	-2.9341	-4.3863	-4.7224	-2.8833	-8.4237	-9.9799

Note: this stable shows MAIC statistic for each model in the row compared to that in the column, see equation (52). We have assumed that $c \approx 0.1$ as a rule of thumb and use that $T = 252$.

Figure 4: Plot of the News Impact curves

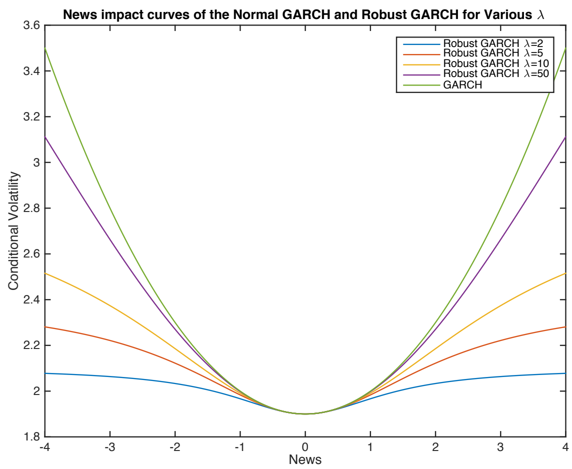


Table 5: Parameter estimates and standards error for the Normal GARCH(1,1) model

Parameter	Microsoft		Yahoo		GE		Intel		Nokia	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
ω	0.1839	0.0206	0.6605	0.0630	0.0298	0.0072	0.0726	0.0190	0.5179	0.0492
α	0.1773	0.0137	0.2641	0.0205	0.0888	0.0109	0.0601	0.0097	0.2046	0.0159
β	0.7598	0.0170	0.7358	0.0122	0.9022	0.0123	0.9174	0.0143	0.7593	0.0154
LL	-5404.1		-9215.7		-5239		-6495.4		-9224.6	

Table 6: Parameter estimates and standards error for the Robust GARCH(1,1)-t model

Parameter	Microsoft		Yahoo		GE		Intel		Nokia	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
ω	0.0348	0.0303	0.3089	0.5829	0.0270	0.0463	0.0603	0.3095	0.4992	0.1423
α	0.1465	0.0328	0.2660	0.2271	0.1348	0.0851	0.1054	0.2893	0.3275	0.0482
β	0.8534	0.0501	0.7339	0.3380	0.8651	0.1028	0.8945	0.3560	0.6724	0.0495
λ	5.7902	0.4593	3.7949	0.4190	4.5198	0.0178	6.4319	0.7924	3.0738	0.2650
LL	-5402		-6843.9		-5409.6		-6098.7		-7305.6	

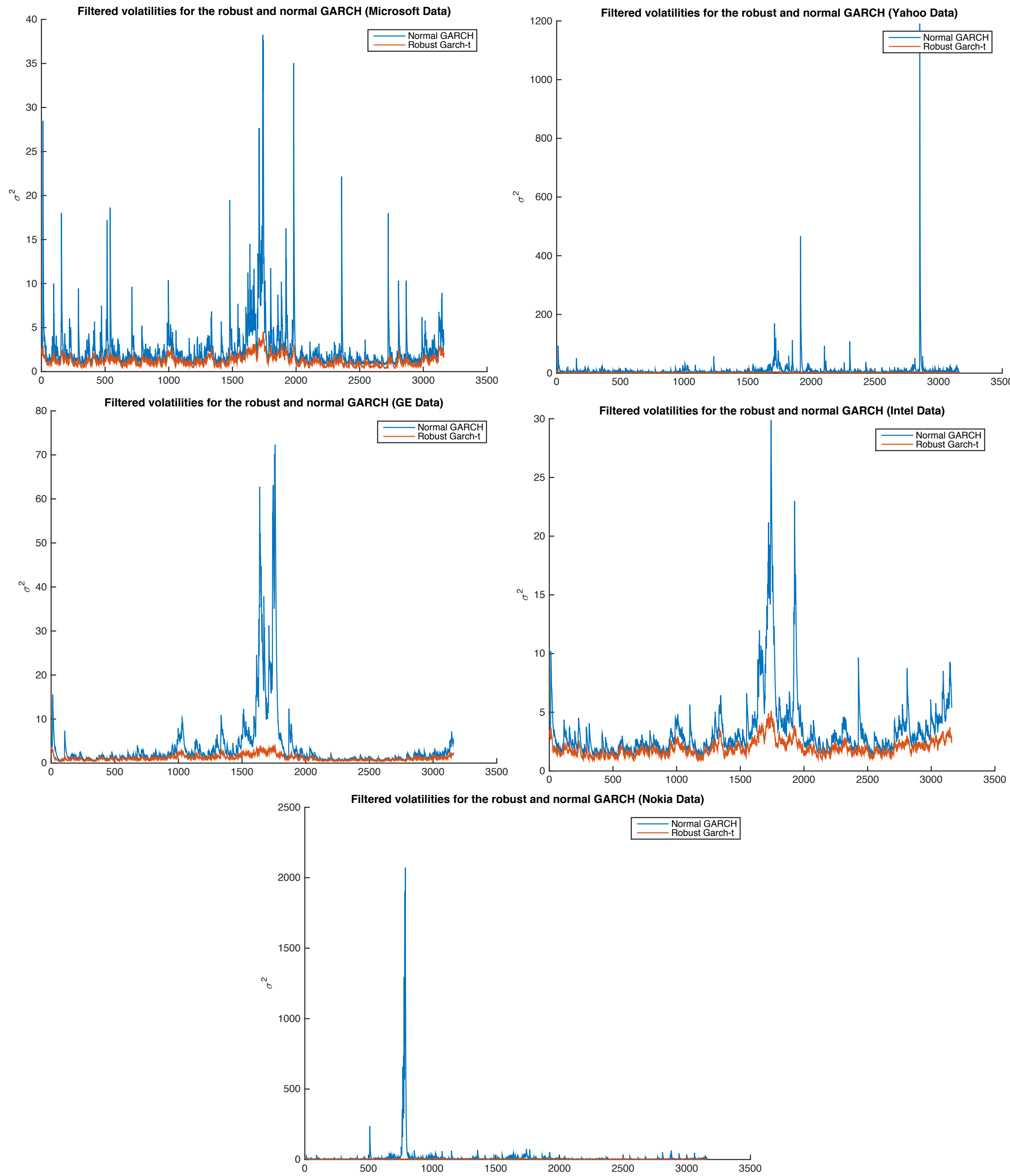
Table 7: 1%, and 5%, VaR for Robust GARCH(1,1)-t model

Parameter	Microsoft		Yahoo		GE		Intel		Nokia	
	1% VaR	5% VaR	1% VaR	5% VaR	1% VaR	5% VaR	1% VaR	5% VaR	1% VaR	5% VaR
Normal GARCH	-4.095	-2.895	-4.951	-3.501	-4.963	-3.509	-5.420	-3.832	-6.462	-4.569
Robust GARCH-t	-3.421	-2.103	-3.734	-2.094	-3.741	-2.193	-3.979	-2.486	-4.698	-2.458

Table 8: MAIC Statistics where we take $L_T^1(\boldsymbol{X_T}, \hat{\boldsymbol{\theta}}_T^1)$ to be the Robust GARCH and $L_T^2(\boldsymbol{X_T}, \hat{\boldsymbol{\theta}}_T^2)$ the normal GARCH.

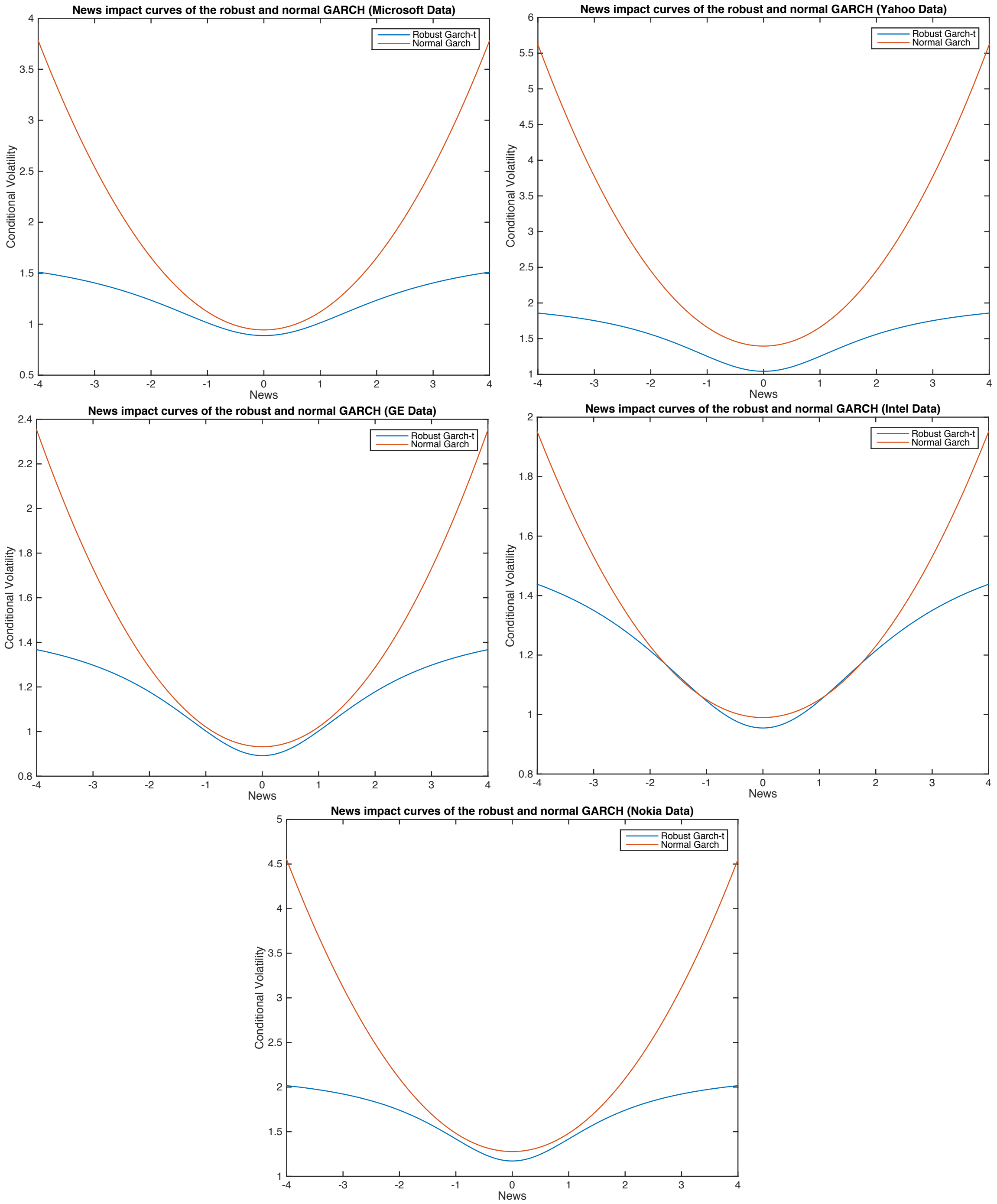
Parameter	Microsoft	Yahoo	GE	Intel	Nokia
Rob. GARCH minus GARCH	-6.00	2363.70	-298.92	388.54	1910.90

Figure 5: Plot of the filtered volatilities



Note: In this figure it is clearly visible how the Normal GARCH overreacts to outliers whereas the Robust GARCH does not, especially for Nokia and Yahoo it is visible that the Normal GARCH overestimates the volatility in an extreme manner this was also visibile from the parameter estimates.

Figure 6: Plot of the News Impact curves



Note: In all the NIC plots below it is visible that the Robust GARCH is indeed more robust to past news than the GARCH.